

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:30:05 ; Search time 8.84706 Seconds

(without alignments)
276.623 Million cell updates/sec

Title: US-10-004-381-25

Perfect score: 258
Sequence: 1 MDEKTGMGSHVVEGLAGE.....LEHHPGQREPMWSGGCKLGG 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60.5	23.4	177	1 HSLV_RHIL0	098C88 rhizobium 1
2	59.5	23.1	342	1 YDEY_ECOLI	P77672 escherichia
3	57.5	22.3	1065	1 KDGJ_HUMAN	075912 homo sapien
4	57	22.1	474	1 TRPG_CORGL	P05650 corynebacte
5	56.5	21.9	215	1 GPMB_SALTI	0820C4 salmonella
6	56.5	21.9	215	1 GPMB_SALTY	082J08 salmonella
7	56.5	21.9	404	1 KVBB_HUMAN	043448 homo sapien
8	56.5	21.9	415	1 BPSJ_BOVIN	028177 bos taurus
9	56.5	21.9	1560	1 SMCK_HUMAN	P41229 homo sapien
10	56	21.7	261	1 YEVE_YEAST	P40078 saccharomyc
11	56	21.7	299	1 YJ52_STRCO	092513 streptomyce
12	55.5	21.5	188	1 HSLV_CAVCR	09A239 caulobacter
13	55.5	21.5	288	1 AOX4_ASRPE	09Yd34 aeropyrum p
14	55.5	21.5	465	1 FXD3_MOUSE	061060 mus musculu
15	54	20.9	376	1 ALGC_STRCO	09IKU4 streptomyce
16	53.5	20.7	352	1 GCP_TPRPA	083686 treponema p
17	53.5	20.7	404	1 KVBB_RAT	063494 rattus norv
18	53	20.5	348	1 PLSX_STNY3	P73950 synochocyst
19	53	20.5	358	1 AROB_ASRPE	09Yej9 aeropyrum p
20	53	20.5	381	1 PTK3_MOUSE	P48759 mus musculu
21	53	20.5	483	1 MURE_CHLCV	0823N2 chlamydomph
22	53	20.5	528	1 WR42_ARATH	09Xex3 arabidopsis
23	53	20.5	589	1 IFEB_ASCSU	P23731 ascaris suu
24	52.5	20.3	184	1 HSLV_BRUME	08Yej1 bruceella me
25	52.5	20.3	184	1 HSLV_BRUSU	08YV11 bruceella su
26	52.5	20.3	344	1 PLSX_SYNEL	08dk15 synecococc
27	52.5	20.3	461	1 SX10_CHICK	09W757 gallus gall
28	52.5	20.3	591	1 VATA_STRP3	08K871 streptococc
29	52	20.2	113	1 XRE_BACSU	P23789 bacillus su
30	52	20.2	270	1 BDNF_CYPCA	090322 cyprinus ca
31	52	20.2	356	1 GBR2_CAEEL	020636 caenorhabdi
32	52	20.2	378	1 HMAN_DROME	P02833 drosophila
33	52	20.2	449	1 AROF_STRCO	P80574 streptomyce

34	52	20.2	499	1 AINX_HUMAN	016352 homo sapien
35	52	20.2	504	1 AINX_MOUSE	P46660 mus musculu
36	52	20.2	505	1 AINX_RAT	P23565 rattus norv
37	52	20.2	2337	1 TOR2_SCHPO	09Y7K2 schizosacch
38	51.5	20.0	181	1 COBU_ECOLI	P46886 escherichia
39	51.5	20.0	273	1 GS80_BACSU	P46878 bacillus su
40	51.5	20.0	323	1 CG96_HUMAN	09Y3A4 homo sapien
41	51.5	20.0	392	1 PRD2_STRGR	P52321 streptomyce
42	51.5	20.0	401	1 KVBI_MOUSE	063277 mus musculu
43	51.5	20.0	408	1 KVBI_MOUSE	028528 mustela put
44	51.5	20.0	419	1 KVBI_HUMAN	014722 homo sapien
45	51.5	20.0	419	1 KVBI_RABIT	09Xt31 oryctolagus

ALIGNMENTS

RESULT 1	HSIV_RHIL0	STANDARD	PRT	177 AA.
ID	HSIV_RHIL0			
AC	098CT8;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	ATP-dependent protease hslv (EC 3.4.25.-)			
GN	HSIV OR M15507.			
OS	Rhizobium loti (Mesorhizobium loti).			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
OC	Phyllobacteriaceae; Mesorhizobium.			
OX	NCBI_TaxId=381;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NAFF03099;			
RX	MEDLINE=21082930; PubMed=11214968;			
RA	Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,			
RA	Matanabe A., Idekawa K., Ishikawa A., Kawashima K., Kimura T.,			
RA	Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,			
RA	Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,			
RA	Takeuchi C., Yamada M., Tabata S.;			
RT	"Complete genome structure of the nitrogen-fixing symbiotic bacterium			
RT	Mesorhizobium loti."			
RL	DNA Res. 7:331-338 (2000).			
CC	-1- FUNCTION: Protease subunit of a proteasome-like degradation			
CC	complex (By similarity).			
CC	-1- SUBUNIT: A double ring-shaped homohexameric of hslv is capped on			
CC	each side by a ring-shaped hslu homohexameric (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).			
CC	-1- SIMILARITY: Belongs to peptidase family T1B. Hslv subfamily.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib.ch).			
CC	-----			
CC	EMBL; AP003005; BAB51533.1; -			
CC	DR MEROPS; T01.006; -			
CC	DR HAMAP; MF_00248; -; 1.			
CC	DR InterPro; IPR001353; Peptidase_T1.			
CC	DR Pfam; PF00227; proteasome; 1.			
CC	DR HydroLase; Protease; Threonine protease; Complete proteome.			
CC	FT ACT SITE			
CC	BY SIMILARITY.			
CC	SEQUENCE 177 AA; 18986 MW; 62F3D3DC6ACB5F05 CRC64;			
CC	-----			
CC	Query Match	23.4%;	Score 60.5;	DB 1; Length 177;
CC	Best local Similarity	41.9%;	Pred. No. 1.7;	
CC	Matches 13; Conservative	6;	Mismatches	5; Indels 7; Gaps 1;
CC	-----			
CC	9 RGHVVEGLAGE-----LEQLRLLEHHP 32			
CC	:			
CC	42 KGVNVIAGFAGATADFTLRLERLAKLQYF 72			

```

RESULT 2
YDEY ECOLI STANDARD; PRT; 342 AA.
ID YDEY ECOLI
AC P77672;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical ABC transporter permease protein ydeY.
GN YDEY OR B1514.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubram S.,
RA Tagami H., Takeda U., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horinuchi T.;
RA "A 570-bp DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RT DNA Res. 3:363-377(1996).
RL
CC -!- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM EGO/YDEY. PROBABLY RESPONSIBLE FOR THE TRANSLOCATION OF
CC THE SUBSTRATE ACROSS THE MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: Belongs to the binding-protein-dependent transport
CC system permease family. AraH/rbcs subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AE000249; AAC74587.1; -.
CC EMBL; D90794; BAA15201.1; -.
CC PIR; E64905; E64905.
DR Ecogene; E613807; ydeY.
DR InterPro; IPR001851; Bac_inmem_transp.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF02653; BPD_transp_2; 1.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 14 34
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 115 135 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 213 233 POTENTIAL.
FT TRANSMEM 252 272 POTENTIAL.
FT TRANSMEM 284 304 POTENTIAL.

```

```

SQ SEQUENCE 342 AA; 36394 MW; 419E505026ABDE33 CRC64;
Query Match 23.1%; Score 59.5; DB 1; Length 342;
Best Local Similarity 35.0%; Pred. No. 4.7;
Matches 14; Conservative 5; Mismatches 10; Indels 11; Gaps 1;
QY 8 WRGHHVHGLAGLEQLPARLEHHHPQGRPEWMSGSGCLTG 47
DB 135 WTGGKWRIGLPAELKQLSA-----PULGVSATIG 163

RESULT 3
KOG1 HUMAN STANDARD; PRT; 1065 AA.
ID KOG1 HUMAN
AC 075812; Q9NZ49;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Diacylglycerol kinase, iota (EC 2.7.1.107) (Diglyceride kinase) (DGK-
DE iota) (DAG kinase iota).
GN DGK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=99047655; PubMed=9830018;
RA Ding L., Traer E., McIntyre T.M., Zimmerman G.A., Prescott S.M.;
RT "The cloning and characterization of a novel human diacylglycerol
RT kinase, DGK-iota";
RL J. Biol. Chem. 273:32746-32752(1998).
RN [2]
RP SEQUENCE OF 135-1065 FROM N.A., AND VARIANT PHE-153.
RX MEDLINE=20173854; PubMed=10706894;
RA Bowse S.J., Sullivan L.S., Ding L., Traer E., Prescott S.M.,
RA Birch D.G., Kemman A., Humphries P., Daiger S.P.;
RT "Evolution of human diacylglycerol kinase iota, DGK, a homolog of
RT Drosophila rdgA, in inherited retinopathy mapping to 7q.";
RL Mol. Vision 6:6-9(2000).
CC -!- CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol = ADP + 1,2-
CC diacylglycerol 3-phosphate.
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -!- SIMILARITY: Belongs to the eukaryotic diacylglycerol kinase
CC family.
CC -!- SIMILARITY: Contains 2 zinc-dependent phospho-ester and DAG
CC binding domains.
CC -!- SIMILARITY: Contains 2 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF061936; AAC62010.1; -.
CC EMBL; AF219939; AAF43006.1; -.
CC EMBL; AF219907; AAF43006.1; JOINED.
CC EMBL; AF219908; AAF43006.1; JOINED.
CC EMBL; AF219909; AAF43006.1; JOINED.
CC EMBL; AF219910; AAF43006.1; JOINED.
CC EMBL; AF219911; AAF43006.1; JOINED.
CC EMBL; AF219912; AAF43006.1; JOINED.
CC EMBL; AF219913; AAF43006.1; JOINED.
CC EMBL; AF219914; AAF43006.1; JOINED.
CC EMBL; AF219915; AAF43006.1; JOINED.
CC EMBL; AF219916; AAF43006.1; JOINED.
CC EMBL; AF219917; AAF43006.1; JOINED.
CC EMBL; AF219918; AAF43006.1; JOINED.
CC EMBL; AF219919; AAF43006.1; JOINED.

```

DR EMBL; AF219920; AAF43006.1; JOINED.
 DR EMBL; AF219921; AAF43006.1; JOINED.
 DR EMBL; AF219922; AAF43006.1; JOINED.
 DR EMBL; AF219923; AAF43006.1; JOINED.
 DR EMBL; AF219924; AAF43006.1; JOINED.
 DR EMBL; AF219925; AAF43006.1; JOINED.
 DR EMBL; AF219926; AAF43006.1; JOINED.
 DR EMBL; AF219927; AAF43006.1; JOINED.
 DR EMBL; AF219928; AAF43006.1; JOINED.
 DR EMBL; AF219929; AAF43006.1; JOINED.
 DR EMBL; AF219930; AAF43006.1; JOINED.
 DR EMBL; AF219931; AAF43006.1; JOINED.
 DR EMBL; AF219932; AAF43006.1; JOINED.
 DR EMBL; AF219933; AAF43006.1; JOINED.
 DR EMBL; AF219934; AAF43006.1; JOINED.
 DR EMBL; AF219935; AAF43006.1; JOINED.
 DR EMBL; AF219936; AAF43006.1; JOINED.
 DR EMBL; AF219937; AAF43006.1; JOINED.
 DR EMBL; AF219938; AAF43006.1; JOINED.
 DR Genew; HGNC:2855; DGKI.
 DR MIM; 604072; .
 DR CO; GO:0005737; C:cytoplasm; TAS.
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0004143; F:diacylglycerol kinase activity; TAS.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR002219; DAG_PE-bind.
 DR InterPro; IPR000756; DAGKA.
 DR InterPro; IPR001206; DAGKC.
 DR Pfam; PF00023; ank; 2.
 DR Pfam; PF00609; DAGKA; 1.
 DR Pfam; PF00781; DAGKC; 1.
 DR ProDom; PD002939; DAGKA; 1.
 DR ProDom; PD005043; DAGKC; 1.
 DR SMART; SM00248; ANK; 2.
 DR SMART; SM00109; C1; 2.
 DR SMART; SM00045; DAGKA; 1.
 DR SMART; SM00046; DAGKC; 1.
 DR PROSITE; PSS0088; ANK_REPEAT; 2.
 DR PROSITE; PSS0297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PSS0479; DAG_PE_BIND_DOM_1; FALSE_NEG.
 DR PROSITE; PSS0081; DAG_PE_BIND_DOM_2; FALSE_NEG.
 DR Transfaser; Kinase; ANK repeat; Repeat; Nuclear protein;
 KM Multigene family; Polymorphism.
 FT DOMAIN 178 232
 FT DOMAIN 251 309 PHORBOL-ESTER AND DAG BINDING 1.
 FT DOMAIN 374 500 PHORBOL-ESTER AND DAG BINDING 2.
 FT DOMAIN 526 683 CATALYTIC-A (POTENTIAL).
 FT REPEAT 958 990 CATALYTIC-B (POTENTIAL).
 FT REPEAT 997 1026 ANK 1.
 FT DOMAIN 20 31 ANK 2.
 FT DOMAIN 69 74 POLY-SER.
 FT DOMAIN 95 102 POLY-ALA.
 FT VARIANT 153 153 L -> F.
 FT VARIANT 153 153 /PTIDSVAR 010190.
 FT CONFLICT 160 160 A -> P (IN REF. 2).
 SQ SEQUENCE 1065 AA; 11696 MM; B84971AA630A799 CRC64;

Query Match 22.3%; Score 57.5; DB 1; Length 1065;
 Best Local Similarity 25.9%; Pred. No. 28;
 Matches 14; Conservative 9; Mismatches 20; Indels 11; Gaps 1;

DB 4 KTTGWRGSH-----VVEGLAGELEQLARLEHHHGGQREPMMSGGCKL 46
 469 KTLNMGGGYDEPVSKITLCVEDGTIVQLDKRNILHVERNDLPPEBLDGVCKL 522

RESULT 4
 TRPC CORGL STANDARD; PRT; 474 AA.
 AC P06560;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Tryptophan biosynthesis protein trpC [includes: indole-3-glycerol
 phosphate synthase (EC 4.1.1.48) (IGPS); N-(5'-phospho-
 ribosyl)anthranilate isomerase (EC 5.3.1.24) (PRAI)].
 GN TRPC OR CGJ3033.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxID=1718;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87117512; PubMed=3808947;
 RA Matsumi K., Sano K., Ohtsubo E.;
 RT "Complete nucleotide and deduced amino acid sequences of the
 RT Brevibacterium lactofermentum tryptophan operon.";
 RL Nucleic Acids Res. 14:10113-10114(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RT Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
 RL FUNCTION: Bifunctional enzyme that catalyzes two sequential steps
 of tryptophan biosynthetic pathway. The first reaction is
 catalyzed by the isomerase, coded by the trpC domain; the second
 reaction is catalyzed by the synthase, coded by the trpC domain.
 CC -1 CATALYTIC ACTIVITY: N-(5-phospho-beta-D-riboseyl)-anthranilate = 1-
 CC (2-carboxyphenylamino)-1-deoxy-D-ribose 5-phosphate
 CC -1 CATALYTIC ACTIVITY: 1-(2-carboxyphenylamino)-1-deoxy-D-ribose 5-
 CC phosphate = 1-(indol-3-yl)glycerol 3-phosphate + CO(2) + H(2)O.
 CC -1 PATHWAY: Tryptophan biosynthesis; third step.
 CC -1 PATHWAY: Tryptophan biosynthesis; fourth step.
 CC -1 SUBUNIT: Monomer.
 CC -1 SIMILARITY: In the N-terminal section; belongs to the trpC family.
 CC -1 SIMILARITY: In the C-terminal section; belongs to the trpC family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X04960; CAA28626.1; .
 DR EMBL; AP005283; BAC00427.1; .
 DR F1R; E24723; E24723.
 DR HSP; P00909; IPT1.
 DR HAMAP; MF_00134; fused; 1.
 DR HAMAP; MF_00135; fused; 1.
 DR InterPro; IPR001468; IGPS.
 DR InterPro; IPR001240; PRAI.
 DR Pfam; PF00218; IGPS; 1.
 DR Pfam; PF00697; PRAI; 1.
 DR ProDom; PD001511; IGPS; 1.
 DR PROSITE; PSS0614; IGPS; 1.
 KM Tryptophan biosynthesis; Isomerase; Lyase; Multifunctional enzyme;
 KM Decarboxylase; Complete; Proteome.
 FT DOMAIN 1 263
 FT DOMAIN 263 474
 FT CONFLICT 88 89
 FT CONFLICT 110 110 A -> G (IN REF. 1).
 FT CONFLICT 130 131 HA -> RP (IN REF. 1).
 FT CONFLICT 153 153 A -> D (IN REF. 1).
 FT CONFLICT 302 302 L -> S (IN REF. 1).
 FT CONFLICT 343 343 D -> G (IN REF. 1).
 FT CONFLICT 381 383 MISSING (IN REF. 1).
 FT CONFLICT 454 474 AGAKDAGLKRTFATISTFHY -> GWGRCRAAENFRDH
 FT CONFLICT 474 474 LHIPLKY (IN REF. 1).
 SQ SEQUENCE 474 AA; 50477 MM; C347C7016B97F9A CRC64;

Query Match 22.1%; Score 57; DB 1; Length 474;
 Best Local Similarity 55.6%; Pred. No. 14;

Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 13 VYVGLAGLEQLRAREH 30
:|||||:|||||:
Db 13 IVEGRGHLEIRARIAH 30

RESULT 5

GPMB_SALTY STANDARD; PRT; 215 AA.

AC 08Z0T4;
DT 28-FEB-2003 (Rel. 41, Last Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable phosphoglycerate mutase gpmb (EC 5.4.2.1)
DE (Phosphoglyceromutase) (PGAM).
GN GPMB OR ST4932 OR T4624.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxId=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RC MEDLINE=21534947; PubMed=11677608;
RA Churchill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Parkhill J., Mungall K.L., Bentley S.D., Holden M.T.G., Sebakhia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Fellwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhimurium CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RC MEDLINE=22531367; PubMed=12644504;
RA Deng W., Lion S.-R., Plunkett G., III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomes of Salmonella enterica serovar Typhimurium Ty2
and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -1- CATALYTIC ACTIVITY: 2-phospho-D-glycerate + 2,3-diphosphoglycerate
= 3-phospho-D-glycerate + 2,3-diphosphoglycerate.
CC -1- PATHWAY: Glycolysis.
CC -1- SIMILARITY: Belongs to the phosphoglycerate mutase family. Gpmb
subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL627284; CAD03416.1; -;
DR EMBL; AE016849; AA072055.1; -;
DR HAMAP; MF_01040; -; 1.
DR InterPro; IPR001345; PG/BPGM_mutase.
DR Pfam; PF00300; PGAM; 1.
DR PROSITE; PS00175; PG_MUTASE; 1.
KW Isomerase; Glycolysis; Complete proteome.
FT ACT_SITE 9 TELB-PHOSPHOINSTITIDINE INTERMEDIATE
FT ACT_SITE 9 (BY SIMILARITY).
FT ACT_SITE 58 58 REQUIRED FOR BINDING CARBOXYL GROUP OF
FT ACT_SITE 150 150 PHOSPHOGLYCERATES (BY SIMILARITY).
FT ACT_SITE BY SIMILARITY.
SQ SEQUENCE 215 AA; 23958 MW; BC47D395F40F819 CRC64;

Query Match 21.9%; Score 56.5; DB 1; Length 215;
Best Local Similarity 26.7%; Pred. No. 6.8;
Matches 16; Conservative 8; Mismatches 23; Indels 13; Gaps 1;

QY 1 MDEKTTGR-----GGHVVGLAGLEQLRAREHHPGGRBPMMSGGCKLG 47
:|||||:|||||:|||||:|||||:
Db 96 LTBEEGRKROLVNGTQDGRIPGSGMQELSDRYAALASCLERPGSRPLVSHGIALG 155

RESULT 6

GPMB_SALTY STANDARD; PRT; 215 AA.

AC 08Z0T8;
DT 28-FEB-2003 (Rel. 41, Last Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable phosphoglycerate mutase gpmb (EC 5.4.2.1)
DE (Phosphoglyceromutase) (PGAM).
GN GPMB OR STW4585.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxId=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RC MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Speeth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
CC -1- CATALYTIC ACTIVITY: 2-phospho-D-glycerate + 2,3-diphosphoglycerate
= 3-phospho-D-glycerate + 2,3-diphosphoglycerate.
CC -1- PATHWAY: Glycolysis.
CC -1- SIMILARITY: Belongs to the phosphoglycerate mutase family. Gpmb
subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AE008915; AAL23400.1; -;
DR StGene; SG27272; gpmb.
DR HAMAP; MF_01040; -; 1.
DR InterPro; IPR001345; PG/BPGM_mutase.
DR Pfam; PF00300; PGAM; 1.
DR PROSITE; PS00175; PG_MUTASE; 1.
KW Isomerase; Glycolysis; Complete proteome.
FT ACT_SITE 9 TELB-PHOSPHOINSTITIDINE INTERMEDIATE
FT ACT_SITE 9 (BY SIMILARITY).
FT ACT_SITE 58 58 REQUIRED FOR BINDING CARBOXYL GROUP OF
FT ACT_SITE 150 150 PHOSPHOGLYCERATES (BY SIMILARITY).
FT ACT_SITE BY SIMILARITY.
SQ SEQUENCE 215 AA; 23868 MW; ADFD335F50FF809 CRC64;

Query Match 21.9%; Score 56.5; DB 1; Length 215;
Best Local Similarity 26.7%; Pred. No. 6.8;
Matches 16; Conservative 8; Mismatches 23; Indels 13; Gaps 1;

QY 1 MDEKTTGR-----GGHVVGLAGLEQLRAREHHPGGRBPMMSGGCKLG 47
:|||||:|||||:|||||:|||||:
Db 96 LTBEEGRKROLVNGTQDGRIPGSGMQELSDRYAALASCLERPGSRPLVSHGIALG 155

RESULT 7

ID	KV3B_HUMAN	STANDARD;	PRT;	404 AA.
AC	O43448;			
DT	16-OCT-2001 (Rel. 40, Created)			
DR	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Voltage-gated potassium channel beta-3 subunit (K+ channel beta-3 subunit) (Kv-beta-3).			
CN	KCNAB3 OR KCNA3B.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteria; Primates; Catarrhini; Hominidae; Homo.			
RX	NCBI_TaxID=9606;			
RN	[1]			
RC	SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.			
RP	TISSUE=Brain;			
RX	MEDLINE=99074289; PubMed=9857044;			
RA	Leichter T.; Baehring R.; Isbrandt D.; Pongs O.;			
RT	"Coexpression of the KCNA3B gene product with Kv1.5 leads to a novel A-type potassium channel.";			
RL	J. Biol. Chem. 273:35095-35101(1998).			
CC	-I- FUNCTION: Accessory potassium channel protein which modulates the activity of the pore-forming alpha subunit. Alters the functional properties of Kv1.5.			
CC	-I- SUBUNIT: Forms heteromultimeric complex with alpha subunits.			
CC	-I- SUBCELLULAR LOCATION: Cytoplasmic (Potential).			
CC	-I- TISSUE SPECIFICITY: Brain-specific expression. Most prominent expression in cerebellum. Weaker signals detected in cortex, occipital lobe, frontal lobe and temporal lobe. Not detected in spinal cord, heart, lung, liver, kidney, pancreas, placenta and skeletal muscle.			
CC	-I- DOMAIN: Alteration of functional properties of alpha subunit is mediated through N-terminal domain of beta subunit (Probable).			
CC	-I- SIMILARITY: Belongs to the shaker potassium channel beta subunit family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by, and for commercial entities requires a license agreement (See http://www.isb-sdb.ch/announce/ or send an email to license@isb-sdb.ch).			
CC	EMBL; AF016411; AAB92499.1; --			
DR	GeneW; HGNC:6230; KCNA3B.			
DR	MIM; 604111; --			
DR	GO; GO:0015455; F:potassium channel regulator activity; TAS.			
DR	GO; GO:0006813; P:potassium ion transport; TAS.			
DR	InterPro; IPR001395; Aldo/Ket red.			
DR	InterPro; IPR005402; KCNA3 Channel.			
DR	InterPro; IPR005399; KCNA3 subunit.			
DR	InterPro; IPR005983; KCNA3 core.			
DR	Pfam; PF00248; Aldo_ket_red_1.			
DR	PRINTS; PR01580; KCNA3CHANNEL.			
DR	PRINTS; PR01577; KCNA3CHANNEL.			
DR	ProDom; PD000288; Aldo/ket_red_2.			
DR	TIGRFAMs; TIGR01293; Kv_beta; 1.			
KW	Ionic channel; ion transport; Potassium transport;			
KW	Voltage-gated channel.			
SQ	SEQUENCE 404 AA; 43530 MW; 08265CC07929A1BA CRC64;			
QY	Query Match 21.9%; Score 56.5; DB 1; Length 404;			
BEST	Local Similarity 31.1%; Pred.No.13;			
MATCHES	14; Conservative 5; Mismatches 3; Indels 23; Gaps 2;			
4	KTGWR-----GG-----HVEGLAGLEQLR 25			
1	: ::			
145	KSNGKRASSIVITTKIFWGQALERTGRSKRTIIIGLGNSLERIQ 169			

RESULT 8
BFS2_BOVIN

ID	BSF3 BOVIN	STANDARD;	PRT;	415 AA.
AC	Q28177;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Phakdin (Beaded filament structural protein 2) (lens fiber cell			
DE	beaded filament protein CP 49) (CP49) (49 kDa cytoskeletal protein)			
DE	(CP 47) (CP47).			
GN	BFSP2.			
OS	Bos taurus (Bovine).			
OC	Eumaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Lens;			
FX	MEDLINE=94075412; PubMed=7504675.			
RA	Merdes A., Gounari F., Georgatos S.D.;			
RT	"The 4.7-kD lens-specific protein phakdin is a tailless intermediate			
RT	filament protein and an assembly partner of filensin.";			
RL	J. Cell Biol. 123:1507-1516 (1993).			
RN	[2]			
RP	REVISIONS.			
RA	Merdes A.;			
RL	Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.			
CC	-1 SUBUNTT: ASSOCIATES WITH BFSP1.			
CC	-1 SUBCELLULAR LOCATION: Membrane- and cytoskeleton-associated.			
CC	-1 TISSUE SPECIFICITY: Lens.			
CC	-1 SIMILARITY: Belongs to the intermediate filament family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X75160; CAA53003.1; -.			
DR	InterPro; IPR001664; IP.			
DR	InterPro; IPR002957; Keratin_I.			
DR	Pfam; PF00038; filament_1.			
DR	PRINTS; PS00148; TYPEKERATIN.			
DR	PROSITE; PS00226; IF; FALSE NEG.			
KW	Intermediate filament; Repeat; Membrane; coiled coil; Cytoskeleton;			
KW	Eye lens protein.			
FT	DOMAIN 1 114 HEAD.			
FT	DOMAIN 115 395 ROD.			
FT	DOMAIN 396 415 TAIL.			
FT	DOMAIN 115 144 COILED COIL (POTENTIAL).			
FT	DOMAIN 199 248 COILED COIL (POTENTIAL).			
FT	DOMAIN 295 395 COILED COIL (POTENTIAL).			
SE	SEQUENCE 415 AA; 45949 MW; F3996852696156 CRC64;			
Query Match	21.9%; Score 56.5; DB 1; Length 415;			
Best Local Similarity	41.7%; Pred. No. 14;			
Matches	15; Conservativity 5; Mismatches 15; Indels 1; Gaps 1;			
Oy	11 GHVTEGLAGLEQLRLLEHHPGQSRPMMSGCKL 46			
Dd	364 GAVVSRLEAEIRERARAEAOOLA-RHLLSHKCOL 398			
RESULT 9				
ID	SMCX HUMAN	STANDARD;	PRT;	1560 AA.
AC	P41229;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Smcx protein (Xel169 protein).			
EN	SMCX OR XE169.			

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fibroblast;
RX MEDLINE=99421443; PubMed=8162017;
RA Wu J., Ellison J., Salido E., Yen P., Mohandas T., Shapiro L.J.;
RT "Isolation and characterization of XE169, a novel human gene that
RL escapes X-inactivation.";
RN Hum. Mol. Genet. 3:153-160(1994).
RN [2]
RP SEQUENCE OF 280-344 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=95038739; PubMed=7951230;
RA Agutink A.I., Mitchell M.J., Mattei M.-G., Borani G., Avner P.A.,
RA Lerner J.L., Bishop C.E.;
RT "A novel X gene with a widely transcribed Y-linked homologue escapes
RT X-inactivation in mouse and human.";
RN Hum. Mol. Genet. 3:879-884(1994).
CC -1- ALTERNATIVE PRODUCTS;
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P41229-1, Sequence=Displayed;
CC Name=2;
CC IsoId=P41229-2, Sequence=VSP_000315;
CC -1- TISSUE SPECIFICITY: Expressed in all tissues examined. Highest
CC levels found in skeletal muscle.
CC -1- MISCELLANEOUS: Escapes X-inactivation.
CC -1- SIMILARITY: Contains 1 ARID domain.
CC -1- SIMILARITY: Contains 1 UMC domain.
CC -1- SIMILARITY: Contains 1 UMD domain.
CC -1- SIMILARITY: Contains 2 PHD-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, L25270, AAA61302.1; -.
DR EMBL, Z29650, CA82758.1; -.
DR PIR, I54361, I54361.
DR Genew; HGNC:11114; SMCX.
DR MIM, 314690; -.
DR InterPro; IPR001606; ARID.
DR InterPro; IPR003347; TF_JmjC.
DR InterPro; IPR003349; TF_JmjN.
DR InterPro; IPR004198; ZnF_CSHC2.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF01388; ARID; 1.
DR Pfam; PF02373; JmjC; 1.
DR Pfam; PF02375; JmjN; 1.
DR Pfam; PF00628; PHD; 2.
DR Pfam; PF02928; zf-CSHC2; 1.
DR SMART; SM00501; BRIGHT; 1.
DR SMART; SM00558; JmjC; 1.
DR SMART; SM00545; JmjN; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS01359; ZF_PHD_1; 2.
DR PROSITE; PS50016; ZF_PHD_2; 1.
KW Zinc-finger; Repeat; Alternative splicing.
FT DOMAIN 13 59 JmjN.
FT FT 76 184 ARID.
FT FT 326 372 PHD-TYPE 1.
FT ZN_FING 501 617 JmjC.
FT ZN_FING 1187 1248 PHD-TYPE 2.
FT VARSPLIC 1370 1372 Missing (in isoform 2).
FT FT /FTID=VSP_000315.
FT C -> Y (IN REF. 2).
FT CONFLICT 342 342

```

```

SQ    SEQUENCE 1560 AA; 175804 MW; 78913F01D0DBCE1 CRC64;
      Query Match          21.9%; Score 56.5; DB 1; Length 1560;
      Best Local Similarity 28.9%; Pred. No. 57;
      Matches 13; Conservative 7; Mismatches 18; Indels 7; Gaps 1;

QY    1 MDEKTTGWRG-----GHVVEGLAGLEQLRLARLENHHPQGGRSP 38
      :|::||::||::||::||::||::||::||::||::||::||::||:
DB    1281 LTERAISWGNRRARQALASBDVTALLGLRAEQRKLQAEPFRPEPP 1325

RESULT 10
YEVE_YEAST STANDARD; PRT; 261 AA.
ID   YEVE_YEAST STANDARD; PRT; 261 AA.
AC   P40078;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DE   10-OCT-2003 (Rel. 42, Last annotation update)
DN   Hypoetical 29.7 kDa protein in RSP5-LCP5 intergenic region.
GN   YER126C OR SGP-ORP47.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxId=4932;
[1]
RP     SEQUENCE FROM N.A.
RC     STRAIN=S288c / AB972;
RX     MEDLINE=97313264; PubMed=9169868;
RA     Dietrich F.S., Mulligan J.T., Hennessey K.M., Yelton M.A., Allen E.,
RA     Arujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA     Cherry J.M., Chung E., Duncan M., Guzman E., Hartwell G.,
RA     Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Laakkari D., Lew H.,
RA     Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA     Oh C., Petrel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA     Smith V., Taylor P., Wei Y., Boettstein D., Davis R.W.;
RT     "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RL     Nature 387:78-81(1997).

CC     -----
CC     This SWISS-PROT entry is copyright. It is produced through a collaboration
CC     between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC     the European Bioinformatics Institute. There are no restrictions on
CC     use by non-profit institutions as long as its content is in no way
CC     modified and this statement is not removed. Usage by and for commercial
CC     entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC     or send an email to license@ebi.ac.uk).
CC     -----
DR   EMBL; U18916; AAC03224.1; -.
DR   PIR; S43218; S43218.
DR   Getmonline; 139205; -.
DR   SGD; S0000928; KR332.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0042273; P:ribosomal large subunit biogenesis; IMP.
DR   InterPro; IPR001047; Ribosomal_S8E.
DR   Pfam; PF01201; Ribosomal_S8e; I.
KW   Hypothetical protein.
SQ    SEQUENCE 261 AA; 29722 MW; 545CBB4250BCB0C9 CRC64;

Query Match          21.7%; Score 56; DB 1; Length 261;
Best Local Similarity 26.7%; Pred. No. 9.7;
Matches 12; Conservative 10; Mismatches 19; Indels 4; Gaps 1;

QY    3 EKTGWGRGHVVEGLAGLEQLRLARLENHHP---GGREPMMSCG 43
      :|::||::||::||::||::||::||::||::||::||::||::||:
DB    40 QKLGWKGMKQPAKKRYAEKVSMRKIKVAHGSKVKGSKPDLDTG 84

RESULT 11
YS52_STRCO STANDARD; PRT; 299 AA.
AC   Q92513;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)

```

DE Hypothetical UPF0042 protein SC01952.
 GN SC01952 OR SCC54.12C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomyces; Streptomycetaceae; Streptomyces.
 RX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2).
 RA Rydning N.J., Ainsa J.A., Hartley N., Bruton C.J., Chater K.F.;
 RT "The whaA sporulation regulatory gene of Streptomyces coelicolor
 A3(2)."
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.T.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wierozorek A., Woodward J., Barrrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2)."
 RL Nature 417:141-147(2002).
 CC -1- SIMILARITY: Belongs to the UPF0042 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AF106003; AAD42167.1; ALT INIT.
 DR EMBL, AL939110; CAB38142.1; ALT_INIT.
 DR PIR, T36025; T36025.
 DR HAMAP; MF_00636; -; 1.
 DR InterPro; IPR005337; UPF0042.
 DR Pfam; PF03668; ATP_bind2; 1.
 KM Hypothetical protein; ATP-binding; Complete proteome.
 FT NP_BIND 23 30 ATP (POTENTIAL).
 SQ SEQUENCE 299 AA; 32821 MW; E69CE8455948B7CD CRC64;
 Query Match 21.7%; Score 56; DB 1; Length 299;
 Best Local Similarity 27.0%; Pred. No. 11;
 Matches 17; Conservative 9; Mismatches 11; Indels 26; Gaps 2;
 QY 11 GHVVEGLAGE-----QLRLRLEHHPGGRP-----NMSSGC 44
 DB 127 GRIVDSIAERELRLRGADLVITSSLVNHELRKADQAGQGEFLRATVWSPF 186
 QY 45 KLG 47
 DB 187 KYG 189
 RESULT 12
 HSLV CAUCR STANDARD; PRT; 188 AA.
 ID HSLV CAUCR
 AC 09A239;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE ATP-dependent protease hslV (EC 3.4.25.-).
 GN HSLV OR CC3727.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;

OC Caulobacteraceae; Caulobacter.
 RX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phade N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Utecherback T., Tran K., Wolf A., Vamathavan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of Caulobacter crescentus."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
 CC complex (By similarity).
 CC -1- SUBUNIT: A double ring-shaped homohexamer of hslV is capped on
 CC each side by a ring-shaped hslU homohexamer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to peptidase family T1b. HslV subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AE006030; AAK25689.1; -;
 DR PIR, B87711; B87711.
 DR HSSP; P31059; 1B94.
 DR MEROPS; T01.006; -;
 DR TIGR; CC3727; -;
 DR HAMAP; MF_00248; -; 1.
 DR InterPro; IPR001353; Peptidase_T1.
 DR Pfam; PF00227; proteasome; 1.
 KM Hydrolyase; Protease; Threonine protease; Complete proteome.
 FT ACT_SITE 14 14 BY SIMILARITY.
 SQ SEQUENCE 188 AA; 19600 MW; ACDDC1FCB10BD61 CRC64;
 Query Match 21.5%; Score 55.5; DB 1; Length 188;
 Best Local Similarity 43.3%; Pred. No. 7.8;
 Matches 13; Conservative 4; Mismatches 6; Indels 7; Gaps 1;
 QY 10 GGVVEGLAGE-----LEQLRLRLEHHP 32
 DB 52 GGVVAGFAGATADAFILIERLEAKLEQYP 81
 RESULT 13
 AOX4 AERPE STANDARD; PRT; 288 AA.
 ID AOX4 AERPE
 AC 09YDX4;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Heme-copper oxidase subunit IV (EC 1.9.3.-).
 GN AOXC OR APE0795.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococciales;
 OC Desulfurococcaceae; Aeropyrum.
 RX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE OF 186-288 FROM N.A.
 RC STRAIN=K1;
 RA Makagi T., Ishikawa R.;
 RT "Heme-copper-oxidase."
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;

```

RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Hatakeyama Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anai A., Kosegi H.,
RA Hosoyama A., Fukui S., Nagai Y., Mshijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudo Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RT DNA Ref. 6:83-101(1999).
CC
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC DR EMBL; AB020482; BAA86073.1; -.
CC DR EMBL; AP000060; BAA79773.1; -.
CC DR PIR; E72671; E72671.
CC KW Oxioreductase; Transmembrane; Respiratory chain; Complete proteome.
CC FT TRANSMEM 205 225 POTENTIAL.
CC FT TRANSMEM 227 247 POTENTIAL.
CC FT TRANSMEM 260 280 POTENTIAL.
CC SQ SEQUENCE 288 AA; 31584 MW; 61B3DDDA610EC381 CRC64;

Query Match 21.5%; Score 55.5; DB 1; Length 288;
Best Local Similarity 28.8%; Pred. No. 12;
Matches 15; Conservative 4; Mismatches 18; Indels 15; Gaps 1;

QY 9 RCGHVVGLAGLEQLRA-----RLHHGQGRPPMGGCK 45
DB 108 RCGHVVGLGCHGSTRATRHGTSTSHNISAPALQRPQGGASTLRSGGR 159

RESULT 14
FXD3_MOUSE STANDARD; PRT; 465 AA.
AC Q61060;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Forward box protein D3 (HNF3/FH transcription factor genesis)
DE (Hepatocyte nuclear factor 3 forkhead homolog 2) (HNF-2).
DE FOXD3 OR HNF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Hromas R.A., Costa R.H., Xu D., Sutton J.L.;
RA Submitted (NOV-1995) to the EMBL/GenBank/DBD databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Labosky P.A., Kaestner K.H.;
RT "The winged helix transcription factor Hth2 is expressed in neural
RT crest and spinal cord during mouse development."
RL Submitted (MAY-1998) to the EMBL/GenBank/DBD databases.
CC -1- FUNCTION: Probable transcription factor.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Contains 1 fork-head domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC
CC DR EMBL; U41047; AAA87569.1; -.
CC DR EMBL; AF067421; AAC28352.1; -.
CC DR HSP; O63245; ZHRH.
CC DR TRANSFAC; T04166; -.
CC DR MGD; MGI:1347473; Foxd3.
CC DR InterPro; IPR001766; TF_Fork_head.
CC DR Pfam; PF00250; Fork_head; 1.
CC DR PRINTS; PR00053; FORKHEAD.
CC DR ProDom; PD000425; TF_Fork_head; 1.
CC DR SMART; SM00339; FH; 1.
CC DR PROSITE; PS00657; FORK_HEAD_1; 1.
CC DR PROSITE; PS00658; FORK_HEAD_2; 1.
CC DR PROSITE; PS50039; FORK_HEAD_3; 1.
CC KW DNA-binding; Nuclear protein; Transcription regulation.
CC FT DOMAIN 106 115 POLY-GLY.
CC FT DNA_BIND 131 225 FORK-HEAD.
CC FT DOMAIN 252 257 POLY-ALA.
CC FT DOMAIN 265 270 POLY-ALA.
CC FT DOMAIN 275 281 POLY-ALA.
CC FT DOMAIN 380 399 POLY-GLY.
CC FT DOMAIN 447 457 POLY-ALA.
CC SQ SEQUENCE 465 AA; 47092 MW; 6F8B5B3D8C7564D CRC64;

Query Match 21.5%; Score 55.5; DB 1; Length 465;
Best Local Similarity 38.6%; Pred. No. 21;
Matches 17; Conservative 3; Mismatches 11; Indels 13; Gaps 2;

QY 7 RCGHVVGLAGLEQLRAL-----EHHPQGRPPMGGC 43
DB 343 GRCGGHHV-----AHQVRAQCAAVVQHRHHRRGIRAPGSGAG 380

RESULT 15
ALC_STRCO STANDARD; PRT; 376 AA.
AC Q9RKU4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative allantoinase (EC 3.5.3.4) (Allantoin amidinohydrolase).
DE SC06248 OR SCMH10.13 OR STAH10.13.
GN Streptomyces coelicolor.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycetaceae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RX Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabdinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)."
RL Nature 417:141-147(2002).
CC -1- CATALYTIC ACTIVITY: Allantoin + H(2)O = (-)-ureidoglycolate +
CC urea.
CC -1- PATHWAY: Degradation of allantoin (purine catabolism); second
CC step.
CC -1- SIMILARITY: Belongs to the allantoinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AL939126; CAB60167.1; -
DR InterPro; IPR005164; Allantoicase.
DR Pfam; PF03561; Allantoicase; 2.
KW Hypothetical protein; Hydrolase; Purine metabolism; Complete proteome.
SQ SEQUENCE 376 AA; 41207 MW; 2247C0B2300CA29 CRC64;

Query Match 20.9%; Score 54; DB 1; Length 376;
Best Local Similarity 44.4%; Pred. No. 26;
Matches 12; Conservative 1; Mismatches 12; Indels 2; Gaps 1;

QY 10 GGHVEGLAGLEEQ--LRARLEHHPOG 34
DB 171 GGHANGFANVAEQRFTHLRKQHPDG 197

Search completed: July 15, 2004, 20:34:28
Job time : 9.84706 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:30:55 ; Search time 36.4941 Seconds
(without alignments)
406.349 Million cell updates/sec

Title: US-10-004-381-25

Perfect score: 258

Sequence: 1 MDEKTTGMRGHHVEGLAGE.....LEHHPGQGRPPMMSGCKLG 47

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

SPTRMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriopl:*
17: sp_archaeopl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73.5	28.5	554	16	Q9A9H1
2	66.5	25.8	739	16	Q82BY9
3	65.5	25.4	189	2	Q8KRD7
4	65.5	25.4	528	16	Q9A365
5	62	24.0	792	16	Q9RDD1
6	61	23.6	429	16	Q9K8S7
7	61	23.6	773	2	Q33541
8	61	23.6	1161	16	Q9L0L0
9	61	23.6	1181	16	Q82DQ5
10	60	23.3	377	16	Q9RKG2
11	60	23.3	420	16	Q8D3W1
12	60	23.3	588	16	Q8DH85
13	59.5	23.1	342	16	Q8KAY8
14	59	22.9	139	4	Q96MD1
15	59	22.9	149	10	Q8L5B8
16	59	22.9	229	16	Q7U554

17	59	22.9	281	10	Q94JG6	Q94JG6 oryza sativ
18	59	22.9	320	16	Q82ZC6	Q82ZC6 enterococcu
19	59	22.9	342	16	Q83RD8	Q83RD8 shigella fl
20	59	22.9	976	5	Q9U158	Q9U158 leishmania
21	59	22.9	1083	3	Q94189	Q94189 cryptococcu
22	59	22.9	1974	5	Q21000	Q21000 caenorhabdi
23	59	22.9	2360	3	Q94188	Q94188 cryptococcu
24	58.5	22.7	274	16	Q8P4G4	Q8P4G4 xanthomonas
25	58.5	22.7	302	16	Q9KRV2	Q9KRV2 vibrio chol
26	58.5	22.7	1206	10	Q7XME8	Q7XME8 oryza sativ
27	58	22.5	333	4	Q8IU04	Q8IU04 homo sapien
28	58	22.5	491	2	Q9ZGDP	Q9ZGDP streptomyce
29	57.5	22.3	303	16	Q8D9M7	Q8D9M7 vibrio vuln
30	57.5	22.3	440	16	Q8K692	Q8K692 streptococc
31	57.5	22.3	442	16	Q879H3	Q879H3 streptococc
32	57.5	22.3	726	5	Q8KX55	Q8KX55 caenorhabdi
33	57.5	22.3	743	5	Q9N313	Q9N313 caenorhabdi
34	57.5	22.3	4001	5	Q8WRQ7	Q8WRQ7 drosophila
35	57.5	22.3	4001	5	Q9VCAB	Q9VCAB drosophila
36	57	22.1	93	10	Q8LME6	Q8LME6 oryza sativ
37	57	22.1	338	2	Q53739	Q53739 streptomyce
38	57	22.1	381	2	Q83YE1	Q83YE1 streptomyce
39	57	22.1	405	2	Q7X372	Q7X372 uncultured
40	57	22.1	473	2	Q9Z405	Q9Z405 pseudomonas
41	57	22.1	487	13	Q8JGR1	Q8JGR1 brachydanio
42	57	22.1	487	13	Q7ZU28	Q7ZU28 brachydanio
43	57	22.1	489	16	Q8FLU7	Q8FLU7 corynebacte
44	57	22.1	495	2	Q50469	Q50469 mycobacteri
45	57	22.1	496	16	P96284	P96284 mycobacteri

ALIGNMENTS

RESULT 1	Q9A9H1	PRELIMINARY:	PRT:	554 AA.
AC	Q9A9H1	01-JUN-2001 (TRMBLrel. 17, Created)		
DT	01-JUN-2001 (TRMBLrel. 17, Last sequence update)			
DT	01-JUN-2003 (TRMBLrel. 24, Last annotation update)			
DE	Hypothetical protein CCI017.			
GN	CCI017.			
OS	Caulobacter crescentus.			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;			
OC	Caulobacteraceae; Caulobacter.			
NCBI_Taxid=155892;				
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 19089 / CB15;			
RX	MEDLINE=21173698; PubMed=11259647;			
RA	Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,			
RA	Bisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,			
RA	Potocka I., Nelson W.C., Newton A., Stephens C., Madock N.D., Ely B.,			
RA	DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,			
RA	Kolman J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,			
RA	Usterback T., Tran K., Wolf A., Vamathevan J., Ermlaeva M., White O.,			
RA	Salberg S.L., Venter U.C., Shapiro L., Fraser C.M.;			
RT	"Complete genome sequence of Caulobacter crescentus."			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).			
DR	EMBL; AE005780; AAK23001.1; -			
DR	PIR; E87375; E87375.			
DR	TIGR; CCI017; -			
DR	GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.			
DR	InterPro; IPR000051; SAM bind.			
KW	Hypothetical protein; Complete proteome.			
SQ	SEQUENCE 554 AA; 61486 MW; 030FF1804A7CDE80 CRC64;			
Query Match	28.5%;	Score 73.5;	DB 16;	Length 554;
Best Local Similarity	35.4%;	Pred. No. 1.4;		
Matches 17; Conservative	6;	Mismatches 20;	Indels 5;	Gaps 1;
QY	5	TTGMRGHHVE-----GLAGELGQLRRLLEHHPGQGRPPMMSGCKLG 47		

Db 45 STEMSAGYTVDMVYTFGYGELNPLRCRLPLTLVGHNAKINACELG 92

RESULT 2

OQ2BY9

ID OQ2BY9 PRELIMINARY; PRT; 739 AA.

AC OQ2BY9

DT 01-JUN-2003 (TREMBlrel. 24, Created)

DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Putative regulatory protein.

GN SAV5565.

OS Streptomyces avermiltilis.

OC Bacteria; Actinobacteria; Actinomycetales;

OC Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=33903;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;

RX MEDLINE=21477403; PubMed=11572948;

RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,

RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,

RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,

RT "Genome sequence of an industrial microorganism Streptomyces

RT avermiltilis: deducing the ability of producing secondary

RT metabolites.";

RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).

RN [12]

RN SEQUENCE FROM N.A.

RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;

RX MEDLINE=22608306; PubMed=12692562;

RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,

RA Sakaki Y., Hattori M., Omura S.;

RT "Complete genome sequence and comparative analysis of the industrial

RT microorganism Streptomyces avermiltilis.";

RT Nat. Biotechnol. 21:526-531(2003).

DR EMBL; AP005043; BAC73277.1; -.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0003824; F:catalytic activity; IEA.

DR InterPro; IPR003594; ATPbind_ATPase.

DR InterPro; IPR001932; PP2C-like.

DR Pfam; PF02518; HATPase_c; 1.

DR SMART; SM00331; PP2C_SIG; 1.

DR Complete proteome.

SQ SEQUENCE 739 AA; 77478 MW; 63853F93EC67A2B6 CRC64;

Query Match 25.8%; Score 66.5; DB 16; Length 739;

Best Local Similarity 32.6%; Pred. No. 14;

Matches 14; Conservative 5; Mismatches 13; Indels 11; Gaps 1;

OY 1 MBEKTTGARG-----GHVVGGLAGLEQLRLRLEHHP 32

Db 341 LDEATGMRGPETGNAPGRLARVWHSERNIEDLRRALRRHP 383

RESULT 3

OQ8RD7

ID OQ8RD7 PRELIMINARY; PRT; 189 AA.

AC OQ8RD7;

DT 01-OCT-2002 (TREMBlrel. 22, Created)

DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Transcription repressor KfrA-like protein.

OS Rhodococcus erythropolis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Nocardiaceae; Rhodococcus.

OX NCBI_TaxID=1833;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MP50;

RX MEDLINE=22083458; PubMed=12089004;

RA Troelt S., Burger S., Calaminius C., Stolz A.;

RT "Cloning and Heterologous Expression of an Enantioselective Amidase

RT from Rhodococcus erythropolis Strain MP50.";

RT Appl. Environ. Microbiol. 68:3279-3286(2002).

DR EMBL; AY026386; AAK11725.1; -.

DR SEQUENCE 189 AA; 20101 MW; 4B8B5D0A093FFFD CRC64;

Query Match 25.4%; Score 65.5; DB 2; Length 189;

Best Local Similarity 45.5%; Pred. No. 4;

Matches 20; Conservative 2; Mismatches 13; Indels 9; Gaps 1;

OY 3 EKTGMRGHHV-----VEGLAGLEQLRLRLEHHPGORE 37

Db 104 EOATERRDGAVTARDRALGVEGLRVELEQRGQLRTARQDARE 147

RESULT 4

OQ9A65

ID OQ9A65 PRELIMINARY; PRT; 528 AA.

AC OQ9A65;

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DE Hypothetical protein CC3341.

GN CC3341.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;

OC Caulobacteraceae; Caulobacter.

OX NCBI_TaxID=155892;

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=ATCC 19089 / CBA15;

RX MEDLINE=21173698; PubMed=11259647;

RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,

RA Eisen J., Heidelberg J.F., Alley M.R.K., Ona N., Maddock J.R.,

RA Potocka I., Nelson W.C., Newton W.C., Stephens C., Phadke N.D., Ely B.,

RA DeBoy R.T., Dodson R.J., Durkin A.S., Gilm M.L., Hatf D.H.,

RA Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,

RA Uterback T., Tran K., Wolf A., Vamathavan J., Emolava M., White O.,

RA Salberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

RT "Complete genome sequence of Caulobacter crescentus.";

RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

DR EMBL; AE005995; AAK25303.1; -.

DR PIR; C87663; C87663.

DR TIGR; CC3341; -.

DR GO; GO:0008168; F:methyltransferase activity; IEA.

DR InterPro; IPR001601; Methyltransf.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 528 AA; 57050 MW; E18A1F4DA1785D5 CRC64;

Query Match 25.4%; Score 65.5; DB 16; Length 528;

Best Local Similarity 37.8%; Pred. No. 12;

Matches 17; Conservative 5; Mismatches 18; Indels 5; Gaps 1;

OY 8 WRGCHVVE-----GLAGLEQLRLRLEHHPGOREPMWSGGCKLG 47

Db 23 MSAGVVEGTGYTGCGYRPLNPVRALPLLAGFAPAKIETACELG 67

RESULT 5

OQ9RDL

ID OQ9RDL PRELIMINARY; PRT; 792 AA.

AC OQ9RDL;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Hypothetical protein SC02560.

GN SC02560 OR SC02560.

OS Streptomyces coelicolor.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RC SEQUENCE FROM N.A.
 RA STRAIN=A3(2);
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleiser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howarth S.,
 RA Huang C.-H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neil S.,
 RA Rabinovitch B., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2)."
 RL Nature 417:141-147(2002).
 DR EMBL: AL939113; CAB6238.1; -
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0003824; F:catalytic activity; IEA.
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR003018; GAP.
 DR InterPro: IPR001932; PP2C-like.
 DR Pfam: PF02518; HATPase_c; 1.
 DR SMART: SM00065; GAP; 1.
 DR SMART: SM00387; HATPase_c; 1.
 DR SMART: SM00331; PP2C_SIG; 1.
 DR Hypoetical protein; Complete proteome.
 KW SEQUENCE 792 AA; 81804 MW; EFWCC6D6685203D CRC64;
 SO
 Query Match 24.0%; Score 62; DB 16; Length 792;
 Best Local Similarity 40.0%; Pred. No. 53;
 Matches 14; Conservative 3; Mismatches 12; Indels 6; Gaps 1;
 QY 8 WRCGHVEGLAGLEQRLARLEHHPOGQREPMMSG 42
 DB 422 WHSGE-----GHLDLRALALEKDPDPMPRMSG 450
 RESULT 6
 O9K8S7 PRELIMINARY; PRT; 429 AA.
 ID O9K8S7
 AC O9K8S7;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Maltose/malodextrin transport system (Maltose/malodextrin-binding
 protein).
 GN BH2926.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 CX NCBI_TaxID=8665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Maeno N.,

RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 halodurans and genomic sequence comparison with Bacillus subtilis.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL: AP001517; BAB06645.1; -
 DR PIR: F84015; F84015.
 DR HSSP: P02928; 1IUD.
 DR GO: GO:0005363; P:maltose transporter activity; IEA.
 DR GO: GO:0005215; P:transporter activity; IEA.
 DR GO: GO:0015768; P:maltose transport; IEA.
 DR GO: GO:0006810; P:transport; IEA.
 DR InterPro: IPR006060; Maltose BP.
 DR InterPro: IPR000437; Prok_lipoprct_S.
 DR InterPro: IPR006059; SBP_Dac1.
 DR InterPro: IPR006061; SBP_dcm1.
 DR Pfam: PF01547; SBP_Dac_1; 1.
 DR PRINTS: PR00181; MALTOSEBP.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 DR PROSITE: PS01037; SBP_BACTERIAL_1; 1.
 KW Complete proteome.
 SO SEQUENCE 429 AA; 46536 MW; 5B17521A0FA19297 CRC64;
 QY 10 GCHVVEGLAGLEQRLARLEHHPOGQREPMMSGCKLG 47
 DB 122 GDVYGLAELPEQLPEQLEGYEGEGADALNVEGAQLG 159
 RESULT 7
 O33541 PRELIMINARY; PRT; 773 AA.
 ID O33541
 AC O33541;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Sensor protein RsaA.
 GN RsaA.
 OS endosymbiont of Rickettsia pachyptila.
 OC Bacteria; Proteobacteria; Gammaproteobacteria;
 OC sulfur-oxidizing symbionts.
 CX NCBI_TaxID=54396;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97438512; PubMed=9293000;
 RA Hughes D.S., Felbeck H., Stein J.L.;
 RT "A histidine protein kinase homolog from the endosymbiont of the
 hydrothermal vent tubeworm Rickettsia pachyptila.";
 RL Appl. Environ. Microbiol. 63:3494-3498(1997).
 CC -1 SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
 KINASES.
 CC EMBL: U93704; AAB71131.1; -
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0005677; F:DNA binding; IEA.
 DR GO: GO:0016301; F:kinase activity; IEA.
 DR GO: GO:0016740; F:transferase activity; IEA.
 DR GO: GO:0000156; F:two-component response regulator activity; IEA.
 DR GO: GO:0000155; F:two-component sensor molecule activity; IEA.
 DR GO: GO:0007600; P:two-component signal transduction system (p. . .); IEA.
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR004358; Bact_sense_pr_C.
 DR InterPro: IPR003467; His_Kinase.
 DR InterPro: IPR003661; His_Kina_N.
 DR InterPro: IPR001789; Response_reg.
 DR Pfam: PF02518; HATPase_c; 1.
 DR Pfam: PF00072; response_reg; 1.
 DR PRINTS: PR00344; BCTR1SENSOR.

DR ProDom; PD000039; Response reg; 1.
 DR SMART; SM00387; HATPase C; 1.
 DR SMART; SM00389; HATPase C; 1.
 DR SMART; SM00448; REC; 1.
 DR PROSITE; PS0109; HIS_KIN; 1.
 DR PROSITE; PS0110; RESPONSE_REGULATOR; 1.
 DR Kinase; Phosphorylation; Sensory transduction; Transferase.
 SQ SEQUENCE 773 AA; 87086 MW; 2C5643B64F373827 CRC64;

Query Match 23.6%; Score 61; DB 2; Length 773;
 Best Local Similarity 43.3%; Pred. No. 68;
 Matches 13; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

8 MKRGHVEGLAGELGRLARLEHHPGORE 37
 346 MKGRRIEELGELGRLARLEHHPGORE 375

RESULT 8
 ID 091010 PRELIMINARY; PRT; 1161 AA.
 AC 091010;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE DNA-directed RNA polymerase beta chain.
 GN RPOB OR SCO4654 OR SCD82.26.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycetaceae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleiser H.M., Denepate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=2196410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.T.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleiser J., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch B., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wierozorek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE +
 CC [RNA] (N).
 CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
 CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1

CC BETA' CHAIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
 CC EMBL; AL939121; CAB77428.1; -.
 DR HSP; Q9K07; IHW.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR InterPro; IPR007121; RNA_pol_B.
 DR InterPro; IPR007644; RNA_pol_Rpb2_1.
 DR InterPro; IPR007642; RNA_pol_Rpb2_2.
 DR InterPro; IPR007645; RNA_pol_Rpb2_3.
 DR InterPro; IPR007120; RNA_pol_Rpb2_6.
 DR InterPro; IPR007641; RNA_pol_Rpb2_7.
 DR Pfam; PF04563; RNA_pol_Rpb2_1; 1.
 DR Pfam; PF04561; RNA_pol_Rpb2_2; 2.
 DR Pfam; PF04565; RNA_pol_Rpb2_3; 1.
 DR Pfam; PF00562; RNA_pol_Rpb2_6; 1.
 DR Pfam; PF04560; RNA_pol_Rpb2_7; 1.
 DR PROSITE; PS01166; RNA_POL_BETA; 1.
 KW DNA-directed RNA polymerase; Transcription; Transferase;
 KW complete proteome.
 SQ SEQUENCE 1161 AA; 128494 MW; A9C27B928E8191A8 CRC64;

Query Match 23.6%; Score 61; DB 16; Length 1161;
 Best Local Similarity 39.5%; Pred. No. 11e+02;
 Matches 15; Conservative 6; Mismatches 15; Indels 2; Gaps 2;

4 KTTGKRGHVEGLAGELGRLARLEHHPGOREPMM 40
 227 KALGWTTEQLIEEF-GEYSMRATLEKDHGTGODDALL 263

RESULT 9
 ID 082D05 PRELIMINARY; PRT; 1181 AA.
 AC 082D05;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative RNA polymerase beta subunit.
 GN RPOB OR SAV4914.
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycetaceae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=1157948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hatori M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 avermitilis: deducing the ability of producing secondary
 metabolites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hatori M., Omura S.;
 RT "Complete genome sequence and comparative analysis of the industrial
 microorganism Streptomyces avermitilis.";
 RL Nat. Biotechnol. 21:526-531(2003).
 CC EMBL; AP005040; BAC72626.1; -.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR InterPro; IPR007121; RNA_pol_B.
 DR InterPro; IPR007644; RNA_pol_Rpb2_1.
 DR InterPro; IPR007642; RNA_pol_Rpb2_2.
 DR InterPro; IPR007645; RNA_pol_Rpb2_3.
 DR InterPro; IPR007120; RNA_pol_Rpb2_6.
 DR InterPro; IPR007641; RNA_pol_Rpb2_7.
 DR Pfam; PF04563; RNA_pol_Rpb2_1; 1.
 DR Pfam; PF04561; RNA_pol_Rpb2_2; 2.
 DR Pfam; PF04565; RNA_pol_Rpb2_3; 1.
 DR Pfam; PF00562; RNA_pol_Rpb2_6; 1.
 DR Pfam; PF04560; RNA_pol_Rpb2_7; 1.
 DR PROSITE; PS01166; RNA_POL_BETA; 1.
 KW DNA-directed RNA polymerase; Transcription; Transferase;
 KW complete proteome.
 SQ SEQUENCE 1161 AA; 128494 MW; A9C27B928E8191A8 CRC64;

DR InterPro: IPR007645; RNA_pol_Rpb2_3.
 DR InterPro: IPR007120; RNA_pol_Rpb2_6.
 DR InterPro: IPR007641; RNA_pol_Rpb2_7.
 DR Pfam: PF04563; RNA_pol_Rpb2_1; 1.
 DR Pfam: PF04561; RNA_pol_Rpb2_2; 1.
 DR Pfam: PF04565; RNA_pol_Rpb2_3; 1.
 DR Pfam: PF00562; RNA_pol_Rpb2_6; 1.
 DR Pfam: PF04560; RNA_pol_Rpb2_7; 1.
 DR PROSITE: PS01166; RNA_POL_BETA; 1.
 DR Complete proteome.
 KM SEQUENCE 1181 AA; 130439 MW; ABBF77ACC879EDB8 CRC64;
 SQ

Query Match 23.6%; Score 61; DB 16; Length 1181;
 Best Local Similarity 39.5%; Pred. No. 1.1e+02;
 Matches 15; Conservative 6; Mismatches 15; Indels 2; Gaps 2;

QY 4 KTTGMRGCHVVEGLAEQLRLRL-HPDGGREPMM 40
 DB 247 KALGWTTEQLLEEF-GEYESMRATLEKDHQGDML 283

RESULT 10
 ID Q9RKG2 PRELIMINARY; PRT; 377 AA.
 AC Q9RKG2;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative isomerase.
 GN SCO3475 OR SCE55.11C OR SCE55.11C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinomycetales;
 OC Streptomyces; Streptomyces.
 NC NCB1_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Murphy L., Harris D.;
 RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Redebach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL M01. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., Batean A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabbittowitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2)."
 RL Nature 417:141-147(2002).
 DR EMBL: AL339116; CAB6179.1; -.
 DR HSSP: P42206; 1BQG.
 DR GO: GO:0016523; F:isomerase activity; IEA.
 DR GO: GO:0008152; P:metabolism; IEA.

DR InterPro: IPR001354; MR_MLE.
 DR Pfam: PF01188; MR_MLE_N; 1.
 DR Pfam: PF02746; MR_MLE_N; 1.
 DR PROSITE: PS00908; MR_MLE_1; 1.
 DR PROSITE: PS00909; MR_MLE_2; 1.
 KM Isomerase; Complete proteome.
 SQ SEQUENCE 377 AA; 40887 MW; 8787D44A30985E91 CRC64;

Query Match 23.3%; Score 60; DB 16; Length 377;
 Best Local Similarity 37.9%; Pred. No. 41;
 Matches 22; Conservative 4; Mismatches 18; Indels 14; Gaps 4;

QY 2 DEKTTGMRGCHVVEGLA-----GELPOLRA-----RLEHHPQGC-REPMMGGCKL 46
 DB 23 DKGVIGM-GEAVVEQARVERMIDLAELVLAGRPRRIENHWQSMYKGFYRGCAVL 79

RESULT 11
 ID Q8D3W1 PRELIMINARY; PRT; 420 AA.
 AC Q8D3W1;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 125, Last annotation update)
 DE ABC-type protease/lipase transport system, ATPase and permease
 DE component.
 GN VV21572.
 OS Vibrio vulnificus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrrio.
 NC NCB1_TaxID=672;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CMCP6;
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
 RA Choy H.E.;
 RT "Complete genome sequence of Vibrio vulnificus CMCP6."
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB016813; AA008435.1; -.
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO: GO:000166; F:nucleotide binding; IEA.
 DR GO: GO:0008233; F:peptidase activity; IEA.
 DR GO: GO:0006810; P:transport; IEA.
 DR InterPro: IPR003593; AAA ATPase.
 DR InterPro: IPR003593; ABC transporter.
 DR Pfam: PF00005; ABC_tran; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
 KM Protease; Complete proteome.
 SQ SEQUENCE 420 AA; 45496 MW; 98A77D0FA3E3A51B CRC64;

Query Match 23.3%; Score 60; DB 16; Length 420;
 Best Local Similarity 21.1%; Pred. No. 46;
 Matches 16; Conservative 13; Mismatches 17; Indels 30; Gaps 3;

QY 1 MDEKTTGMRGCHVVEGLAEQLRLRL-HPDGGREPMM 38
 DB 133 MGOATTTMSASALASAVGRLEISDKVQVDELMNPLPRPRGHFELAVSYNHP-GVNEP 195

QY 39 MMSG-----GCKLG 47
 DB 196 VLSGAQFSPGCAVG 211

RESULT 12
 ID Q8DH85 PRELIMINARY; PRT; 588 AA.
 AC Q8DH85;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

```

DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE Penicillin-binding protein.
GN TLR2074.
OS Synecoccus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecoccus.
OX NCBI_TaxID=32046;
(1)
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneo T., Sato S., Ikeuchi M., Katoh H., Sasa moto S.,
RA Watanabe A., Iiguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005376; BAC09626.1; -.
DR GO; GO:0008658; F:penicillin binding; IEA.
DR GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.
DR InterPro; IPR005311; PBP_dimer.
DR InterPro; IPR001460; Transpeptidase.
DR Pfam; PF03717; PBP_dimer; 1.
DR Pfam; PF00905; Transpeptidase; 1.
DR Complete proteome.
SQ SEQUENCE 588 AA; 65048 MW; 9E24A8544A102719 CRC64;

Query Match 23.3%; Score 60; DB 16; Length 588;
Best Local Similarity 57.1%; Pred. No. 67;
Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 12 HVVEGLAGELEQRLARLEHP 32
DB 442 HVVDGLYDNGQLQKRLHP 462

RESULT 13
Q8XAY8 PRELIMINARY; PRT; 342 AA.
AC Q8XAY8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, last annotation update)
DE Putative transport system permease protein.
GN YDEY OR Z2191 OR ECS2121.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
(1)
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=1120651;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
(2)
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 050952;
RX MEDLINE=21156231; PubMed=11358796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AB005353; AAG56252.1; -.

```

```

DR EMBL; AP002557; BAB35544.1; -.
DR PIR; A90894; A90894.
DR PIR; H85723; H85723.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001851; Bac_inmem_transp.
DR Pfam; PF02653; BPD_transp_2; 1.
DR Complete proteome.
SQ SEQUENCE 342 AA; 36365 MW; 6953A23C07C305A8 CRC64;

Query Match 23.1%; Score 59.5; DB 16; Length 342;
Best Local Similarity 35.0%; Pred. No. 42;
Matches 14; Conservative 5; Mismatches 10; Indels 11; Gaps 1;

QY 8 WRGHHVGLAGELEQRLARLEHPGQREPMSSGCKLG 47
DB 135 WTGGKWIIEGLPAELKQLSA-----PULLGVSAIG 163

RESULT 14
Q96MD1 PRELIMINARY; PRT; 139 AA.
AC Q96MD1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, last annotation update)
DE Hypothetical protein FLJ32575.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
(1)
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashta H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Waga tsuna M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.;
RT "NEBO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK051137; BAB71369.1; -.
DR Hypothetical protein.
SQ SEQUENCE 139 AA; 15014 MW; 3A59D86197B48F19 CRC64;

Query Match 22.9%; Score 59; DB 4; Length 139;
Best Local Similarity 44.8%; Pred. No. 18;
Matches 13; Conservative 3; Mismatches 11; Indels 2; Gaps 1;

QY 3 EKTGWGHHVGLAGELEQRLARLEHP 31
DB 92 ERATGWRLSHVEGOSGLQATSCR--HH 118

RESULT 15
Q8LSB8 PRELIMINARY; PRT; 149 AA.
AC Q8LSB8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
DE P0671D01.7 protein (P0445H04.30 protein).
GN P0671D01.7 OR P0445H04.30.
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrharioideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
(1)
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;

```

RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
 RT clone:P0671D01."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0445H04."
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP003284; BAB91731.1; -
 DR EMBL; AP004364; BAC06318.1; -
 DR Gramene; O815B8; -
 SQ SEQUENCE 149 AA; 16716 MW; 0081630FBE03018D CRC64;

Query Match 22.9%; Score 59; DB 10; Length 149;
 Best Local Similarity 43.3%; Pred. No. 19;
 Matches 13; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 17 LAGLEQLRLRLHHHPQGRPPMMSGGCKL 46
 DB 77 LAGDLSTYRESLRFHPKPPAPAKGGL 106

Search completed: July 15, 2004, 20:35:47
 Job time : 38.4941 secs

is Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:29:44 ; Search time 53.0824 Seconds

(without alignments)
250.172 Million cell updates/sec

Title: US-10-004-381-25

Perfect score: 258
Sequence: 1 MDEKTTGMRGHHVEGLAGE.....LEHHPGQGRPMWSSGCKLG 47

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_290404:*

1: geneseqp1980s:*\n2: geneseqp1990s:*\n3: geneseqp2000s:*\n4: geneseqp2001s:*\n5: geneseqp2002s:*\n6: geneseqp2003as:*\n7: geneseqp2003bs:*\n8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	258	100.0	47	5	ABG67079 Streptavi
2	258	100.0	47	6	ABUS7601 Synthetic
3	243	94.2	47	5	ABG67080 Streptavi
4	243	94.2	47	6	ABUS7602 Streptavi
5	241	93.4	61	5	ABG67078 Streptavi
6	241	93.4	61	6	ABUS7600 Synthetic
7	234	90.7	75	5	ABG67077 Streptavi
8	234	90.7	75	6	ABUS7599 Synthetic
9	227	88.0	89	5	ABG67076 Streptavi
10	227	88.0	89	6	ABUS7598 Synthetic
11	214	82.9	101	5	ABG67075 Streptavi
12	214	82.9	101	6	ABUS7597 Streptavi
13	211	81.8	101	5	ABG67073 Streptavi
14	211	81.8	101	6	ABUS7595 Streptavi
15	208	80.6	38	5	ABG67087 Streptavi
16	208	80.6	38	6	ABUS7609 Streptavi
17	208	80.6	426	5	ABG67091 Maltose-b
18	208	80.6	426	6	ABUS7613 Streptavi
19	208	80.6	479	5	ABG67089 Maltose-b
20	208	80.6	479	6	ABUS7611 Streptavi
21	199	77.1	39	5	ABG67081 Streptavi
22	199	77.1	39	6	ABUS7603 Synthetic
23	152	58.9	29	5	ABG67083 Streptavi
24	152	58.9	29	6	ABUS7605 Synthetic
25	133	51.6	89	5	ABG67082 Streptavi

26	133	51.6	89	6	ABUS7604 Synthetic
27	94	36.4	44	5	ABG67086 Streptavi
28	88	34.1	44	6	ABUS7608 Streptavi
29	70	27.1	102	5	ABG67065 Streptavi
30	70	27.1	102	6	ABUS7587 Streptavi
31	68	26.4	101	5	ABG67066 Streptavi
32	68	26.4	101	6	ABUS7588 Streptavi
33	63	24.4	145	5	ABP09375 Human ORF
34	62	24.0	164	4	ABG63519 Human gas
35	62	24.0	336	6	ABG67778 Phototrab
36	61	23.6	288	6	ABR96152 Human NOV
37	60.5	23.4	741	4	ABG25060 Novel hum
38	59.5	23.1	84	5	ABP34332 Human hel
39	59.5	23.1	342	3	ABR15903 E. coli p
40	59.5	23.1	342	6	ABU14680 Protein e
41	59.5	23.1	538	4	ABG25106 Novel hum
42	59	22.9	139	6	ADA55387 Human pro
43	59	22.9	140	2	AAW40203 FRAP homo
44	59	22.9	147	4	AAU19547 Peptide #
45	59	22.9	147	4	ABR39099 Peptide #

ALIGNMENTS

RESULT 1	ABG67079 standard, peptide, 47 AA.
ID	ABG67079
XX	AC ABG67079;
XX	DT 24-SEP-2002 (first entry)
XX	DE Streptavidin-binding peptide SB19 truncation mutant #5.
XX	KW Streptavidin-binding peptide; mutant; mutein.
OS	Escherichia coli.
PN	WO200238580-A1.
XX	PD 16-MAY-2002.
XX	PF 31-OCT-2000; 2000WO-US041717.
XX	PR 31-OCT-2000; 2000WO-US041717.
XX	PA (GENO) GEN HOSPITAL CORP.
PI	Szostrak JM, Wilson DS, Keefe AD;
XX	WPI; 2002-500155/53.
XX	Novel peptide with high affinity for streptavidin, is expressed as part of fusion protein to facilitate detection, quantitation and purification of desired protein.
XX	Claim 9; Fig 5; 56pp; English.
PS	The invention describes a peptide (I) which binds streptavidin with a dissociation constant less than 10 mM or 23 nM, where the amino acid sequence of (I) does not contain an HPQ, HPM, HPN or HQP motif, is not disulfide bonded or cyclised. A fusion protein comprising a protein of interest covalently linked to (I) is useful for purifying a desired protein from a sample by contacting the sample with streptavidin under conditions that allow complex formation between the fusion protein and the streptavidin, isolating the complex and recovering the sample. The protein, and thus purifying the desired protein from the sample. The invention describes a method for producing a streptavidin-binding fusion protein. This sequence represents a truncation mutant of the streptavidin-binding peptide SB-19 isolated from Escherichia coli using a method described in the specification

SQ Sequence 47 AA;

Query Match 100.0%; Score 258; DB 5; Length 47;
Best Local Similarity 100.0%; Pred. No. 2,3e-28;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRLEHHHPQGRPPMWSGGCKLG 47
DB 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRLEHHHPQGRPPMWSGGCKLG 47

RESULT 2

ABU57601

ID ABU57601 standard; peptide; 47 AA.

AC ABU57601;

DT 09-APR-2003 (first entry)

DE Synthetic peptide SB19 C-terminal deletion mutant, C4.

KM Streptavidin-binding peptide; mRNA display; peptide library;
fusion protein; mutant; mutein; SB19.

XX Synthetic.

OS US2002155578-A1.

PN 24-OCT-2002.

PF 31-OCT-2001; 2001US-00004381.

PR 31-OCT-2000; 2000US-0244541P.

PA (SZOS/) SZOSTAK J W.

PA (WILS/) WILSON D S.

PA (KEEF/) KEEFE A D.

PI Szostak JW, Wilson DS, Keefe AD;

DR WPI; 2003-182639/18.

Novel peptide which binds streptavidin with high affinity useful as
affinity tags for purifying fusion proteins containing proteins of
interest.

Claim 14; Fig 5; 36pp; English.

The invention relates to a peptide which binds streptavidin with a
dissociation constant less than 10 micro M, where the peptide is not
disulphide bonded or cyclised, or with a dissociation constant less than
10 micro M, where the amino acid sequence of the peptide does not contain
an HPO, HPN or HQP motif. The peptides are generated by the mRNA
display method to create a library of peptides. Also include are a
nucleic acid encoding the peptide, a vector comprising the nucleic acid,
a fusion protein (and its encoding nucleic acid/vector) comprising a
protein of interest covalently linked to the streptavidin binding peptide
and detecting the presence of the fusion protein in a sample using
streptavidin or streptavidin-containing compounds. The fusion protein
(which comprises the desired protein of interest covalently linked to the
streptavidin binding peptide) is useful for purifying a protein of
interest from a sample, which involves expressing the fusion protein in
the sample, contacting the sample with streptavidin under conditions that
allow complex formation between the fusion protein and the streptavidin,
isolating the complex and recovering the fusion protein, thereby
purifying the protein of interest from the sample. The present sequence
is a mutated version of SB19, a streptavidin binding peptide of the
invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
deletions and a truncated N-terminal peptide with point mutations

SQ Sequence 47 AA;

Query Match 100.0%; Score 258; DB 6; Length 47;

Best Local Similarity 100.0%; Pred. No. 2,3e-28;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRLEHHHPQGRPPMWSGGCKLG 47
DB 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRLEHHHPQGRPPMWSGGCKLG 47

RESULT 3

ABG67080

ID ABG67080 standard; peptide; 47 AA.

AC ABG67080;

DT 24-SEP-2002 (first entry)

DE Streptavidin-binding peptide SB19 truncation mutant #6.

KM Streptavidin-binding peptide; mutant; mutein.

OS Escherichia coli.

FN WO200238580-A1.

PD 16-MAY-2002.

PF 31-OCT-2000; 2000WO-US0411717.

PR 31-OCT-2000; 2000WO-US0411717.

PA (GEHO) GEN HOSPITAL CORP.

PI Szostak JW, Wilson DS, Keefe AD;

DR WPI; 2002-500155/53.

Novel peptide with high affinity for streptavidin, is expressed as part
of fusion protein to facilitate detection, quantitation and purification
of desired protein.

Claim 9; Fig 5; 56pp; English.

The invention describes a peptide (I) which binds streptavidin with a
dissociation constant less than 10 mM or 23 nM, where the amino acid
sequence of (I) does not contain an HPO, HPN or HQP motif, is not
disulphide bonded or cyclised. A fusion protein comprising a protein of
interest covalently linked to (I) is useful for purifying a desired
protein from a sample by contacting the sample with streptavidin under
conditions that allow complex formation between the fusion protein and
the streptavidin, isolating the complex and recovering the fusion
protein, and thus purifying the desired protein from the sample. The
invention describes a method for producing a streptavidin-binding fusion
protein. This sequence represents a truncation mutant of the streptavidin
-binding peptide SB-19 isolated from Escherichia coli using a method
described in the specification

SQ Sequence 47 AA;

Query Match 94.2%; Score 243; DB 5; Length 47;
Best Local Similarity 95.7%; Pred. No. 2,8e-26;
Matches 45; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRLEHHHPQGRPPMWSGGCKLG 47
DB 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRLEHHHPQGRPPMWSGGCKLG 47

RESULT 4

ABU57602

ID ABU57602 standard; peptide; 47 AA.

AC ABU57602;

XX

DT 09-APR-2003 (first entry)
 XX Streptavidin-binding synthetic peptide SB19 point mutant M1.
 DE Streptavidin-binding peptide; mRNA display; peptide library;
 XX fusion protein; mutant; mutein; SB19.
 KW
 XX Synthetic.
 OS
 XX US2002155578-A1.
 PN
 XX 24-OCT-2002.
 PD
 XX 31-OCT-2001; 2001US-00004381.
 PF
 XX 31-OCT-2000; 2000US-0244541P.
 PR
 XX (SZOS/) SZOSTAK J W.
 XX (WILS/) WILSON D S.
 PA (KEEF/) KEEFE A D.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;
 XX
 XX WPI; 2003-182639/18.
 DR
 XX
 PT Novel peptide which binds streptavidin with high affinity useful as
 PT affinity tags for purifying fusion proteins containing proteins of
 PT interest.
 XX
 XX Claim 14; Fig 5; 36pp; English.
 PS
 XX The invention relates to a peptide which binds streptavidin with a
 CC dissociation constant less than 10 micro M, where the peptide is not
 CC disulfide bonded or cyclised, or with a dissociation constant less than
 CC 10 micro M, where the amino acid sequence of the peptide does not contain
 CC an HQQ, HPW, or HQP motif. The peptides are generated by the mRNA
 CC display method to create a library of peptides. Also include are a
 CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
 CC a fusion protein (and its encoding nucleic acid/vector) comprising a
 CC protein of interest covalently linked to the streptavidin binding peptide
 CC and detecting the presence of the fusion protein in a sample using
 CC streptavidin or streptavidin-containing compounds. The fusion protein
 CC (which comprises the desired protein of interest covalently linked to the
 CC streptavidin binding peptide) is useful for purifying a protein of
 CC interest from a sample, which involves expressing the fusion protein in
 CC the sample, contacting the sample with streptavidin under conditions that
 CC allow complex formation between the fusion protein and the streptavidin,
 CC isolating the complex and recovering the fusion protein, thereby
 CC purifying the protein of interest from the sample. The present sequence
 CC is a mutated version of SB19, a streptavidin binding peptide of the
 CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
 CC deletions and a truncated N-terminal peptide with point mutations
 CC
 XX
 SQ Sequence 47 AA;
 Query Match 94.2%; Score 243; DB 6; Length 47;
 Best Local Similarity 95.7%; Pred. No. 2.8e-26;
 Matches 45; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVVEGLAGELRRLRLEHHPQGRPPMSSGGCKLG 47
 |||||
 DB 1 MDEKTTGMRGHHVVEGLAGELRRLRLEHHPQGRPPMSSGGCKLG 47

RESULT 5
 ABG67078
 ID ABG67078 standard; peptide; 61 AA.
 XX
 AC ABG67078;
 XX
 DT 24-SBP-2002 (first entry)
 XX
 DE Streptavidin-binding peptide SB19 truncation mutant #4.

XX
 KW Streptavidin-binding peptide; mutant; mutein.
 XX Escherichia coli.
 OS
 XX WO200238580-A1.
 PN
 XX 16-MAY-2002.
 PD
 XX 31-OCT-2000; 2000WO-US041717.
 PF
 XX 31-OCT-2000; 2000WO-US041717.
 PR
 XX (GEHO) GEN HOSPITAL CORP.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;
 XX
 XX WPI; 2002-500155/53.
 DR
 XX
 PT Novel peptide with high affinity for streptavidin, is expressed as part
 PT of fusion protein to facilitate detection, quantitation and purification
 PT of desired protein.
 XX
 XX Claim 9; Fig 5; 56pp; English.
 PS
 XX The invention describes a peptide (I) which binds streptavidin with a
 CC dissociation constant less than 10 mM or 23 nM, where the amino acid
 CC sequence of (I) does not contain an HQQ, HPW, HPN or HQP motif, is not
 CC disulfide bonded or cyclised. A fusion protein comprising a protein of
 CC interest covalently linked to (I) is useful for purifying a desired
 CC protein from a sample by contacting the sample with streptavidin under
 CC conditions that allow complex formation between the fusion protein and
 CC the streptavidin, isolating the complex and recovering the fusion
 CC protein, and thus purifying the desired protein from the sample. The
 CC invention describes a method for producing a streptavidin-binding fusion
 CC protein. This sequence represents a truncation mutant of the streptavidin
 CC -binding peptide SB-19 isolated from Escherichia coli using a method
 CC described in the specification
 CC
 XX
 SQ Sequence 61 AA;
 Query Match 93.4%; Score 241; DB 5; Length 61;
 Best Local Similarity 77.0%; Pred. No. 7.3e-26;
 Matches 47; Conservative 0; Mismatches 0; Indels 14; Gaps 1;

QY 1 MDEKTTGMRGHHVVEGLAGELRRLRLEHHPQGRPPMSSGGCKLG 46
 |||||
 DB 1 MDEKTTGMRGHHVVEGLAGELRRLRLEHHPQGRPPMSSGGCKLG 60

QY 47 G 47
 DB 61 G 61

RESULT 6
 ABUS7600
 ID ABUS7600 standard; peptide; 61 AA.
 XX
 AC ABUS7600;
 XX
 DT 09-APR-2003 (first entry)
 XX
 DE Synthetic peptide SB19 C-terminal deletion mutant, C3.
 XX
 KW Streptavidin-binding peptide; mRNA display; peptide library;
 KW fusion protein; mutant; mutein; SB19.
 XX
 OS Synthetic.
 XX
 XX US2002155578-A1.
 XX
 PN
 XX 24-OCT-2002.
 PD
 XX

PF 31-OCT-2001; 2001US-00004381.
 XX
 PR 31-OCT-2000; 2000US-0244541P.
 XX
 PA (SZOS/) SZOSTAK J W.
 PA (WILS/) WILSON D S.
 PA (KEEF/) KEEFE A D.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;
 XX
 DR WPI; 2003-182639/18.
 XX
 PT Novel peptide which binds streptavidin with high affinity useful as
 PT affinity tags for purifying fusion proteins containing proteins of
 PT interest.
 XX
 PS Claim 14; Fig 5; 36pp; English.

CC The invention relates to a peptide which binds streptavidin with a
 CC dissociation constant less than 10 micro M, where the peptide is not
 CC disulfide bonded or cyclised, or with a dissociation constant less than
 CC 10 micro M, where the amino acid sequence of the peptide does not contain
 CC an HPO, HPM, HPN or HOP motif. The peptides are generated by the mRNA
 CC display method to create a library of peptides. Also include are a
 CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
 CC a fusion protein (and its encoding nucleic acid/vector) comprising a
 CC protein of interest covalently linked to the streptavidin binding peptide
 CC and detecting the presence of the fusion protein in a sample using
 CC streptavidin or streptavidin-containing compounds. The fusion protein
 CC (which comprises the desired protein of interest covalently linked to the
 CC streptavidin binding peptide) is useful for purifying a protein of
 CC interest from a sample, which involves expressing the fusion protein in
 CC the sample, contacting the sample with streptavidin under conditions that
 CC allow complex formation between the fusion protein and the streptavidin,
 CC isolating the complex and recovering the fusion protein, thereby
 CC purifying the protein of interest from the sample. The present sequence
 CC is a mutated version of SB19, a streptavidin binding peptide of the
 CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
 CC deletions and a truncated N-terminal peptide with point mutations
 XX
 SQ Sequence 61 AA;

Query Match 93.4%; Score 241; DB 6; Length 61;
 Best Local Similarity 77.0%; Pred. No. 7.3e-26;
 Matches 47; Conservative 0; Mismatches 0; Indels 14; Gaps 1;

QY 1 MDEKTTGMRGSHVVEGLAEQLRLRLHHHPQGOREP-----WMSGGCKL 46
 DB 1 MDEKTTGMRGSHVVEGLAEQLRLRLHHHPQGOREPVEVEDVDGLVQMSGGCKL 60
 QY 47 G 47
 DB 61 G 61

RESULT 7
 ABG67077
 ID ABG67077 standard; peptide; 75 AA.
 XX
 AC ABG67077;
 XX
 DT 24-SBP-2002 (first entry)
 XX
 DE Streptavidin-binding peptide SB19 truncation mutant #3.
 XX
 KW Streptavidin-binding peptide; mutant; mutein.
 XX
 OS Escherichia coli.
 XX
 PN WO200238580-A1.
 XX
 PD 16-MAY-2002.
 XX

PF 31-OCT-2000; 2000WO-US041717.
 XX
 PR 31-OCT-2000; 2000WO-US041717.
 XX
 PA (GENO) GEN HOSPITAL CORP.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;
 XX
 DR WPI; 2002-500155/53.
 XX
 PT Novel peptide with high affinity for streptavidin, is expressed as part
 PT of fusion protein to facilitate detection, quantitation and purification
 PT of desired protein.
 XX
 PS Claim 9; Fig 5; 56pp; English.

CC The invention describes a peptide (1) which binds streptavidin with a
 CC dissociation constant less than 10 nM or 23 nM, where the amino acid
 CC sequence of (1) does not contain an HPO, HPM, HPN or HOP motif, is not
 CC disulfide bonded or cyclised. A fusion protein comprising a protein of
 CC interest covalently linked to (1) is useful for purifying a desired
 CC protein from a sample by contacting the sample with streptavidin under
 CC conditions that allow complex formation between the fusion protein and
 CC the streptavidin, isolating the complex and recovering the fusion
 CC protein, and thus purifying the desired protein from the sample. The
 CC invention describes a method for producing a streptavidin-binding fusion
 CC protein. This sequence represents a truncation mutant of the streptavidin
 CC -binding peptide SB-19 isolated from Escherichia coli using a method
 CC described in the specification
 XX
 SQ Sequence 75 AA;

Query Match 90.7%; Score 234; DB 5; Length 75;
 Best Local Similarity 62.7%; Pred. No. 8.7e-25;
 Matches 47; Conservative 0; Mismatches 0; Indels 28; Gaps 1;

QY 1 MDEKTTGMRGSHVVEGLAEQLRLRLHHHPQGOREP----- 38
 DB 1 MDEKTTGMRGSHVVEGLAEQLRLRLHHHPQGOREPVEVEDVDGLHGVVAG 60

QY 39 -----WMSGGCKLG 47
 DB 61 ILDPVEMMSGCKLG 75

RESULT 8
 ABUS7599
 ID ABUS7599 standard; peptide; 75 AA.
 XX
 AC ABUS7599;
 XX
 DT 09-APR-2003 (first entry)
 XX
 DE Synthetic peptide SB19 C-terminal deletion mutant, C2.
 XX
 KW Streptavidin-binding peptide; mRNA display; peptide library;
 KW fusion protein; mutant; mutein; SB19.
 XX
 OS Synthetic.
 XX
 PN US2002155578-A1.
 XX
 PD 24-OCT-2002.
 XX
 PF 31-OCT-2001; 2001US-00004381.
 XX
 PR 31-OCT-2000; 2000US-0244541P.
 XX
 PA (SZOS/) SZOSTAK J W.
 PA (WILS/) WILSON D S.
 PA (KEEF/) KEEFE A D.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;

XX DR WPI; 2003-182639/18.
 XX PT Novel peptide which binds streptavidin with high affinity useful as
 PT affinity tags for purifying fusion proteins containing proteins of
 PT interest.
 XX PS Claim 14; Fig 5; 36pp; English.
 XX CC The invention relates to a peptide which binds streptavidin with a
 CC dissociation constant less than 10 micro M, where the peptide is not
 CC disulfide bonded or cyclised, or with a dissociation constant less than
 CC 10 micro M, where the amino acid sequence of the peptide does not contain
 CC an HQQ, HPW, HPN or HQP motif. The peptides are generated by the RNA
 CC display method to create a library of peptides. Also include are a
 CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
 CC a fusion protein (and its encoding nucleic acid/vector) comprising a
 CC protein of interest covalently linked to the streptavidin binding peptide
 CC and detecting the presence of the fusion protein in a sample using
 CC streptavidin or streptavidin-containing compounds. The fusion protein
 CC (which comprises the desired protein of interest covalently linked to the
 CC streptavidin binding peptide) is useful for purifying a protein of
 CC interest from a sample, which involves expressing the fusion protein in
 CC the sample, contacting the sample with streptavidin under conditions that
 CC allow complex formation between the fusion protein and the streptavidin,
 CC isolating the complex and recovering the fusion protein, thereby
 CC purifying the protein of interest from the sample. The present sequence
 CC is a mutated version of SB19, a streptavidin binding peptide of the
 CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
 CC deletions and a truncated N-terminal peptide with point mutations
 XX SQ Sequence 75 AA:
 Query Match 90.7%; Score 234; DB 6; Length 75;
 Best Local Similarity 62.7%; Pred. No. 8.7e-25;
 Matches 47; Conservative 0; Mismatches 0; Indels 28; Gaps 1;
 QY 1 MDEKTTGMRGSHVVEGLAGELEQLRARLEHHPGQREPP----- 38
 DB 1 MDEKTTGMRGSHVVEGLAGELEQLRARLEHHPGQREPLVQVEDVDEGLVODLHGAVG 60
 QY 39 -----MMSGGCKLG 47
 DB 61 LDDPVEVMMSGCKLG 75
 Db
 QY
 Db
 AC
 XX
 XX ABG67076; standard; peptide; 89 AA.
 XX DT 24-SEP-2002 (first entry)
 XX DE Streptavidin-binding peptide SB19 truncation mutant #2.
 XX KM Streptavidin-binding peptide; mutant; mutein.
 XX OS Escherichia coli.
 XX PN W0200238580-A1.
 XX PD 16-MAY-2002.
 XX PF 31-OCT-2000; 2000MO-US041717.
 XX PR 31-OCT-2000; 2000MO-US041717.
 XX PA (GEHO) GEN HOSPITAL CORP.
 XX PI Szostrak JW, Wilson DS, Keefe AD;
 XX DR WPI; 2002-500155/53.

XX PT Novel peptide with high affinity for streptavidin, is expressed as part
 PT of fusion protein to facilitate detection, quantitation and purification
 PT of desired protein.
 XX PS Claim 9; Fig 5; 56pp; English.
 XX CC The invention describes a peptide (I) which binds streptavidin with a
 CC dissociation constant less than 10 mM or 23 nM, where the amino acid
 CC sequence of (I) does not contain an HQQ, HPW, HPN or HQP motif, is not
 CC disulfide bonded or cyclised. A fusion protein comprising a protein of
 CC interest covalently linked to (I) is useful for purifying a desired
 CC protein from a sample by contacting the sample with streptavidin under
 CC conditions that allow complex formation between the fusion protein and
 CC the streptavidin, isolating the complex and recovering the fusion
 CC protein, and thus purifying the desired protein from the sample. The
 CC invention describes a method for producing a streptavidin-binding fusion
 CC protein. This sequence represents a truncation mutant of the streptavidin
 CC -binding peptide SB-19 isolated from Escherichia coli using a method
 CC described in the specification
 XX SQ Sequence 89 AA:
 Query Match 88.0%; Score 227; DB 5; Length 89;
 Best Local Similarity 52.8%; Pred. No. 1e-23;
 Matches 47; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
 QY 1 MDEKTTGMRGSHVVEGLAGELEQLRARLEHHPGQREPP----- 38
 DB 1 MDEKTTGMRGSHVVEGLAGELEQLRARLEHHPGQREPLVQVEDVDEGLVODLHGAVG 60
 QY 39 -----MMSGGCKLG 47
 DB 61 LDDPVEKLTITWFKKFKKNVSMMSGCKLG 89
 Db
 QY
 Db
 AC
 XX
 XX ABU57598; standard; peptide; 89 AA.
 XX DT 09-APR-2003 (first entry)
 XX DE Synthetic peptide SB19 C-terminal deletion mutant, C1.
 XX KM Streptavidin-binding peptide; mRNA display; peptide library;
 XX fusion protein; mutant; mutein; SB19.
 XX OS Synthetic.
 XX PN US2002155578-A1.
 XX PD 24-OCT-2002.
 XX PF 31-OCT-2001; 2001US-00004381.
 XX PR 31-OCT-2000; 2000US-0244541P.
 XX PA (SZOS/) SZOSTAK J W.
 XX PA (WILS/) WILSON D S.
 XX PA (KEEF/) KEEFE A D.
 XX PI Szostrak JW, Wilson DS, Keefe AD;
 XX DR WPI; 2003-182639/18.
 XX PT Novel peptide which binds streptavidin with high affinity useful as
 PT affinity tags for purifying fusion proteins containing proteins of
 PT interest.
 XX PS Claim 14; Fig 5; 36pp; English.

CC The invention relates to a peptide which binds streptavidin with a
 CC dissociation constant less than 10 micro M, where the peptide is not
 CC disulfide bonded or cyclised, or with a dissociation constant less than
 CC 10 micro M, where the amino acid sequence of the peptide does not contain
 CC an HPO, HPW, HPN or HOP motif. The peptides are generated by the mRNA
 CC display method to create a library of peptides. Also include are a
 CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
 CC a fusion protein (and its encoding nucleic acid/vector) comprising a
 CC protein of interest covalently linked to the streptavidin binding peptide
 CC and detecting the presence of the fusion protein in a sample using
 CC streptavidin or streptavidin-containing compounds. The fusion protein
 CC (which comprises the desired protein of interest covalently linked to the
 CC streptavidin binding peptide) is useful for purifying a protein of
 CC interest from a sample, which involves expressing the fusion protein in
 CC the sample, contacting the sample with streptavidin under conditions that
 CC allow complex formation between the fusion protein and the streptavidin,
 CC isolating the complex and recovering the fusion protein, thereby
 CC purifying the protein of interest from the sample. The present sequence
 CC is a mutated version of SB19, a streptavidin binding peptide of the
 CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
 CC deletions and a truncated N-terminal peptide with point mutations

XX Sequence 89 AA:

Query Match 88.0%; Score 227; DB 6; Length 89;
 Best Local Similarity 52.8%; Pred. No. 1e-23;
 Matches 47; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MDEKTTGMRGGHVVEGLAGLEQLRARLEHHPOGOREP----- 38
 DB 1 MDEKTTGMRGGHVVEGLAGLEQLRARLEHHPOGOREPVEEDVDEGLVODLHGAVAG 60

QY 39 -----MMSGGCKLG 47
 DB 61 LLDPEVKLLTDMFKFKNVSKDCKMTFYLEMYDMMSGGCKLG 89

RESULT 11
 ABG67075
 ID ABG67075 standard; peptide; 101 AA.

AC ABG67075;
 DT 24-SEP-2002 (first entry)

DE Streptavidin-binding peptide SB19 truncation mutant #1.

KW Streptavidin-binding peptide; mutant; mutein.

OS Escherichia coli.

OS Synthetic.

XX WO200238580-A1.

XX 16-MAY-2002.

PF 31-OCT-2000; 2000WO-US041717.

PR 31-OCT-2000; 2000WO-US041717.

XX (GENO) GEN HOSPITAL CORP.

PI Szostak JW, Wilson DS, Keefe AD;

DR WPI; 2002-500155/53.

XX Novel peptide with high affinity for streptavidin, is expressed as part
 PT of fusion protein to facilitate detection, quantitation and purification
 PT of desired protein.

PS Claim 9; Fig 5; 56pp; English.

XX The invention describes a peptide (I) which binds streptavidin with a

CC dissociation constant less than 10 mM or 23 nM, where the amino acid
 CC sequence of (I) does not contain an HPO, HPW, HPN or HOP motif, is not
 CC disulfide bonded or cyclised. A fusion protein comprising a protein of
 CC interest covalently linked to (I) is useful for purifying a desired
 CC protein from a sample by contacting the sample with streptavidin under
 CC conditions that allow complex formation between the fusion protein and
 CC the streptavidin, isolating the complex and recovering the fusion
 CC protein, and thus purifying the desired protein from the sample. The
 CC invention describes a method for producing a streptavidin-binding fusion
 CC protein. This sequence represents a truncation mutant of the streptavidin
 CC -binding peptide SB-19 isolated from Escherichia coli using a method
 CC described in the specification

XX Sequence 101 AA;

Query Match 82.9%; Score 214; DB 5; Length 101;
 Best Local Similarity 44.6%; Pred. No. 7.7e-22;
 Matches 45; Conservative 2; Mismatches 0; Indels 54; Gaps 1;

QY 1 MDEKTTGMRGGHVVEGLAGLEQLRARLEHHPOGOREP----- 40
 DB 1 MDEKTTGMRGGHVVEGLAGLEQLRARLEHHPOGOREPVEEDVDEGLVODLHGAVAG 60

QY 41 -----SGGCKLG 47
 DB 61 LLDPEVKLLTDMFKFKNVSKDCKMTFYLEMYDMMSGGCKLG 101

RESULT 12
 ABUS7597
 ID ABUS7597 standard; peptide; 101 AA.

AC ABUS7597;

DT 09-APR-2003 (first entry)

DE Streptavidin-binding synthetic peptide SB19 full length peptide.

KW Streptavidin-binding peptide; mRNA display; peptide library;

OS Synthetic.

XX US2002155578-A1.

XX 24-OCT-2002.

PF 31-OCT-2001; 2001US-00004381.

PR 31-OCT-2000; 2000US-0244541P.

XX (SZOS/) SZOSTAK J W.

PA (WILS/) WILSON D S.

PA (KEEF/) KEEFE A D.

PI Szostak JW, Wilson DS, Keefe AD;

DR WPI; 2003-182639/18.

XX Novel peptide which binds streptavidin with high affinity useful as
 PT affinity tags for purifying fusion proteins containing proteins of
 PT interest.

PS Claim 14; Fig 5; 36pp; English.

XX The invention relates to a peptide which binds streptavidin with a
 CC dissociation constant less than 10 micro M, where the peptide is not
 CC disulfide bonded or cyclised, or with a dissociation constant less than
 CC 10 micro M, where the amino acid sequence of the peptide does not contain
 CC an HPO, HPW, HPN or HOP motif. The peptides are generated by the mRNA
 CC display method to create a library of peptides. Also include are a
 CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
 CC a fusion protein (and its encoding nucleic acid/vector) comprising a

CC protein of interest covalently linked to the streptavidin binding peptide
CC and detecting the presence of the fusion protein in a sample using
CC streptavidin or streptavidin-containing compounds. The fusion protein
CC (which comprises the desired protein of interest covalently linked to the
CC streptavidin binding peptide) is useful for purifying a protein of
CC interest from a sample, which involves expressing the fusion protein in
CC the sample, contacting the sample with streptavidin under conditions that
CC allow complex formation between the fusion protein and the streptavidin,
CC isolating the complex and recovering the fusion protein, thereby
CC purifying the protein of interest from the sample. The present sequence
CC is a streptavidin binding peptide of the invention
XX

SQ Sequence 101 AA;

Query Match 82.9%; Score 214; DB 6; Length 101;

Best Local Similarity 44.6%; Pred. No. 7.7e-22;

Matches 45; Conservative 2; Mismatches 0; Indels 54; Gaps 1;

QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREPM-----40

Db 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREPLVQVEVDVDEGLVQDLHGAVAG 60

QY 41 -----SGGCKLG 47

Db 61 LLDPEKLTLDWFKKFNVSCKCKMTFYLEMYDMSGCKLG 101

RESULT 13

ABG67073 standard; peptide; 101 AA.

AC ABG67073;

DT 24-SEP-2002 (first entry)

DE Streptavidin-binding peptide SB19.

KM Streptavidin-binding peptide.

XX Escherichia coli.

PN MO200238580-A1.

PD 16-MAY-2002.

PF 31-OCT-2000; 2000MO-US041717.

PR 31-OCT-2000; 2000MO-US041717.

PA (GEHO) GEN HOSPITAL CORP.

PI Szoetack JW, Wilson DS, Keefe AD;

DR WPI; 2002-500155/53.

PT Novel peptide with high affinity for streptavidin, is expressed as part

PT of fusion protein to facilitate detection, quantitation and purification

PS Claim 9; Fig 3; 56pp; English.

CC The invention describes a peptide (I) which binds streptavidin with a

CC dissociation constant less than 10 mM or 23 nM, where the amino acid

CC sequence of (I) does not contain an HPO, HPW, HPN or HOP motif, is not

CC disulfide bonded or cyclized. A fusion protein comprising a protein of

CC interest covalently linked to (I) is useful for purifying a desired

CC protein from a sample by contacting the sample with streptavidin under

CC conditions that allow complex formation between the fusion protein and

CC the streptavidin, isolating the complex and recovering the fusion

CC protein, and thus purifying the desired protein from the sample. The

CC invention describes a method for producing a streptavidin-binding fusion

CC protein. This sequence represents a streptavidin-binding peptide isolated

CC from Escherichia coli using a method described in the specification

XX SQ Sequence 101 AA;

Query Match 81.8%; Score 211; DB 5; Length 101;

Best Local Similarity 95.0%; Pred. No. 2e-21;

Matches 38; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREPM 40

Db 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREPLV 40

RESULT 14

ABUS7595 standard; peptide; 101 AA.

AC ABUS7595;

DT 09-APR-2003 (first entry)

DE Streptavidin-binding synthetic peptide SB19.

KM Streptavidin-binding peptide; mRNA display; peptide library;

XX fusion protein.

OS Synthetic.

PN US2002155578-A1.

PD 24-OCT-2002.

PF 31-OCT-2001; 2001US-00004381.

PR 31-OCT-2000; 2000US-0244541P.

PA (SZOS/) SZOSTACK J W.

PA (WILS/) WILSON D S.

PA (KEEF/) KEEFE A D.

PI Szoetack JW, Wilson DS, Keefe AD;

DR WPI; 2003-182639/18.

PS Claim 14; Fig 3; 36pp; English.

CC The invention relates to a peptide which binds streptavidin with a

CC dissociation constant less than 10 micro M, where the peptide is not

CC disulfide bonded or cyclized, or with a dissociation constant less than

CC 10 micro M, where the amino acid sequence of the peptide does not contain

CC an HPO, HPW, HPN or HOP motif. The peptides are generated by the mRNA

CC display method to create a library of peptides. Also include are a

CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,

CC a fusion protein (and its encoding nucleic acid/vector) comprising a

CC protein of interest covalently linked to the streptavidin binding peptide

CC and detecting the presence of the fusion protein in a sample using

CC streptavidin or streptavidin-containing compounds. The fusion protein

CC (which comprises the desired protein of interest covalently linked to the

CC streptavidin binding peptide) is useful for purifying a protein of

CC interest from a sample, which involves expressing the fusion protein in

CC the sample, contacting the sample with streptavidin under conditions that

CC allow complex formation between the fusion protein and the streptavidin,

CC isolating the complex and recovering the fusion protein, thereby

CC purifying the protein of interest from the sample. The present sequence

CC is a streptavidin binding peptide of the invention

XX SQ Sequence 101 AA;

Query Match 81.8%; Score 211; DB 6; Length 101;

Best Local Similarity 95.0%; Pred. No. 2e-21;

Matches 38; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVEGLAGLEQLRLRHHHPQGOREPM 40
 |||
 Db 1 MDEKTTGMRGHHVEGLAGLEQLRLRHHHPQGOREPLV 40
 |||

RESULT 15

ABG67087
 ID ABG67087 standard; peptide; 38 AA.

AC ABG67087;

DT 24-SEP-2002 (first entry)

DE Streptavidin-binding peptide and affinity tag.

KW Streptavidin-binding peptide.

OS Synthetic.

PN WO200238580-A1.

PD 16-MAY-2002.

PF 31-OCT-2000; 2000WO-US041717.

PR 31-OCT-2000; 2000WO-US041717.

PA (GENO) GEN HOSPITAL CORP.

PI Szostak JW, Wilson DS, Keefe AD;

DR WPI; 2002-500155/53.

PT Novel peptide with high affinity for streptavidin, is expressed as part
 of fusion protein to facilitate detection, quantitation and purification
 of desired protein.

PS Claim 13; Fig 7A; 56pp; English.

CC The invention describes a peptide (I) which binds streptavidin with a
 CC dissociation constant less than 10 mM or 23 nM, where the amino acid
 CC sequence of (I) does not contain an HPO, HPM, HPN or HOP motif, is not
 CC disulfide bonded or cyclised. A fusion protein comprising a protein of
 CC interest covalently linked to (I) is useful for purifying a desired
 CC protein from a sample by contacting the sample with streptavidin under
 CC conditions that allow complex formation between the fusion protein and
 CC the streptavidin, isolating the complex and recovering the fusion
 CC protein, and thus purifying the desired protein from the sample. The
 CC invention describes a method for producing a streptavidin-binding fusion
 CC protein. This sequence represents a streptavidin-binding peptide used in
 CC the creation of streptavidin-binding DNA-tagged peptide described in the
 CC invention

SQ Sequence 38 AA;

Query Match 80.6%; Score 208; DB 5; Length 38;
 Best Local Similarity 100.0%; Pred. No. 1.7e-21;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVEGLAGLEQLRLRHHHPQGOREP 38
 |||
 Db 1 MDEKTTGMRGHHVEGLAGLEQLRLRHHHPQGOREP 38
 |||

Search completed: July 15, 2004, 20:34:02
 Job time : 55.0824 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:35:51 ; Search time 35.7647 Seconds
(without alignments)
332.096 Million cell updates/sec

Title: US-10-004-381-25_COPY_1_38

Perfect score: 208

Sequence: 1 MDEKTTGMRGHHVBSLAGELEQLRLARLHHHOGQREP 38

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	208	100.0	38	US-10-004-381-35	Sequence 35, Appl
2	208	100.0	47	US-10-004-381-25	Sequence 25, Appl
3	208	100.0	61	US-10-004-381-24	Sequence 24, Appl
4	208	100.0	75	US-10-004-381-23	Sequence 23, Appl
5	208	100.0	89	US-10-004-381-22	Sequence 22, Appl
6	208	100.0	101	US-10-004-381-19	Sequence 19, Appl
7	208	100.0	101	US-10-004-381-21	Sequence 21, Appl
8	208	100.0	426	US-10-004-381-41	Sequence 41, Appl
9	208	100.0	479	US-10-004-381-38	Sequence 38, Appl
10	193	92.8	47	US-10-004-381-26	Sequence 26, Appl
11	149	71.6	39	US-10-004-381-27	Sequence 27, Appl
12	127	61.1	89	US-10-004-381-28	Sequence 28, Appl
13	102	49.0	29	US-10-004-381-29	Sequence 29, Appl
14	67	32.2	102	US-10-004-381-11	Sequence 11, Appl
15	66.5	32.0	739	US-10-156-761-13097	Sequence 13097, A

16	66	31.7	101	13	US-10-004-381-12	Sequence 12, Appl
17	62	29.8	527	12	US-10-425-114-70961	Sequence 70961, A
18	62	29.8	529	12	US-10-425-114-37475	Sequence 37475, A
19	62	29.8	649	12	US-10-425-114-70144	Sequence 70144, A
20	59	28.4	139	15	US-10-094-749-2955	Sequence 2955, Ap
21	59	28.4	140	12	US-09-902-180-6	Sequence 6, Appl1
22	59	28.4	249	12	US-10-112-944-859	Sequence 859, App
23	59	28.4	249	12	US-10-112-944-860	Sequence 860, App
24	59	28.4	249	12	US-10-112-944-861	Sequence 861, App
25	59	28.4	249	12	US-10-112-944-862	Sequence 862, App
26	59	28.4	342	9	US-09-912-020-260	Sequence 260, App
27	59	28.4	342	12	US-10-287-122A-42604	Sequence 42604, A
28	59	28.4	869	16	US-10-437-963-156107	Sequence 156107, A
29	58.5	28.1	175	16	US-10-437-963-147749	Sequence 147749, A
30	58	27.9	288	12	US-10-112-944-410	Sequence 410, App
31	58	27.9	298	12	US-10-423-584-2	Sequence 2, Appl1
32	58	27.9	300	12	US-10-112-944-411	Sequence 411, App
33	58	27.9	306	12	US-10-112-944-413	Sequence 413, App
34	58	27.9	333	12	US-10-112-944-412	Sequence 412, App
35	58	27.9	333	12	US-10-423-582-2	Sequence 2, Appl1
36	58	27.9	333	12	US-10-411-120-69	Sequence 69, Appl
37	58	27.9	333	12	US-10-411-120-97	Sequence 97, Appl
38	58	27.9	1181	14	US-10-156-761-12448	Sequence 12448, A
39	58	27.9	2222	15	US-10-369-493-3923	Sequence 3923, Ap
40	57	27.4	114	12	US-10-425-114-61050	Sequence 61050, A
41	57	27.4	474	9	US-09-738-626-6856	Sequence 6856, Ap
42	56.5	27.2	939	16	US-10-437-963-132193	Sequence 132193, A
43	55.5	26.7	1336	16	US-10-437-963-115864	Sequence 115864, A
44	55.5	26.7	1388	12	US-10-332-089-2	Sequence 2, Appl1
45	55.5	26.7	1388	14	US-10-146-473-82	Sequence 82, Appl

ALIGNMENTS

```

RESULT 1
US-10-004-381-35
; Sequence 35, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-35

Query Match      100.0%; Score 208; DB 13; Length 38;
Best Local Similarity 100.0%; Pred. No. 2.7e-20;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MDEKTTGMRGHHVBSLAGELEQLRLARLHHHOGQREP 38
      |||
Db      1 MDEKTTGMRGHHVBSLAGELEQLRLARLHHHOGQREP 38

RESULT 2
US-10-004-381-25
; Sequence 25, Application US/10004381
; Publication No. US20020155578A1

```

```
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ PRIOR FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 25
/ LENGTH: 47
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-25
```

```
Query Match          100.0%; Score 208; DB 13; Length 47;
Best Local Similarity 100.0%; Pred. No. 3.4e-20;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGOREP 38
DB 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGOREP 38
```

```
RESULT 3
US-10-004-381-24
/ Sequence 24, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ CURRENT FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ PRIOR FILING DATE: 2000-10-31
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 24
/ LENGTH: 61
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-24
```

```
Query Match          100.0%; Score 208; DB 13; Length 61;
Best Local Similarity 100.0%; Pred. No. 4.5e-20;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGOREP 38
DB 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGOREP 38
```

```
RESULT 4
US-10-004-381-23
/ Sequence 23, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
```

```
/ TITLE OF INVENTION: THEREOF
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ PRIOR FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 23
/ LENGTH: 75
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-23
```

```
Query Match          100.0%; Score 208; DB 13; Length 75;
Best Local Similarity 100.0%; Pred. No. 5.6e-20;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGOREP 38
DB 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGOREP 38
```

```
RESULT 5
US-10-004-381-22
/ Sequence 22, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ CURRENT FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ PRIOR FILING DATE: 2000-10-31
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 22
/ LENGTH: 89
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-22
```

```
Query Match          100.0%; Score 208; DB 13; Length 89;
Best Local Similarity 100.0%; Pred. No. 6.7e-20;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGOREP 38
DB 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGOREP 38
```

```
RESULT 6
US-10-004-381-19
/ Sequence 19, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ CURRENT FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
```

```
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-19
```

```
Query Match          100.0%; Score 208; DB 13; Length 101;
Best Local Similarity 100.0%; Pred. No. 7.7e-20;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 38
Db 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 38
```

```
RESULT 7
US-10-004-381-21
; Sequence 21, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: WILSON, JACK W.
; APPLICANT: SZOSTAK, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-21
```

```
Query Match          100.0%; Score 208; DB 13; Length 101;
Best Local Similarity 100.0%; Pred. No. 7.7e-20;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 38
Db 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 38
```

```
RESULT 8
US-10-004-381-41
; Sequence 41, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: WILSON, JACK W.
; APPLICANT: SZOSTAK, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 426
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: designed protein
US-10-004-381-41
```

```
Query Match          100.0%; Score 208; DB 13; Length 426;
Best Local Similarity 100.0%; Pred. No. 3.6e-19;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 38
Db 378 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 415
```

```
RESULT 9
US-10-004-381-38
; Sequence 38, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 479
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Designed protein
US-10-004-381-38
```

```
Query Match          100.0%; Score 208; DB 13; Length 479;
Best Local Similarity 100.0%; Pred. No. 4.1e-19;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 38
Db 378 MDEKTTGMRGHHVVEGLAGLEQLRLARLEHHPOGQREP 415
```

```
RESULT 10
US-10-004-381-26
; Sequence 26, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: WILSON, JACK W.
; APPLICANT: SZOSTAK, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-26
```

Query Match 92.8%; Score 193; DB 13; Length 47;
Best Local Similarity 94.7%; Pred. No. 3.3e-18;
Matches 36; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MDEKTTGMRGHHVEGLAGLEQLRRLRLEHHPOGOREP 38
Db 1 MDEKTTGMRGHHVEGLAGLEQLRRLRLEHHPOGOREP 38

RESULT 11
US-10-004-381-27

; Sequence 27, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:

; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES

; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; CURRENT FILING DATE: 2001-10-31

; PRIOR APPLICATION NUMBER: US 60/244,541
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 39
; TYPE: PRT

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide

US-10-004-381-27

Query Match 71.6%; Score 149; DB 13; Length 39;
Best Local Similarity 100.0%; Pred. No. 1.9e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 GHVVEGLAGLEQLRRLRLEHHPOGOREP 38
Db 3 GHVVEGLAGLEQLRRLRLEHHPOGOREP 30

RESULT 12
US-10-004-381-28

; Sequence 28, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:

; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES

; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; CURRENT FILING DATE: 2001-10-31

; PRIOR APPLICATION NUMBER: US 60/244,541
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 89
; TYPE: PRT

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide

US-10-004-381-28

Query Match 61.1%; Score 127; DB 13; Length 89;
Best Local Similarity 100.0%; Pred. No. 3.9e-09;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 EGLAGLEQLRRLRLEHHPOGOREP 38
Db 3 EGLAGLEQLRRLRLEHHPOGOREP 26

RESULT 13
US-10-004-381-29

; Sequence 29, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:

; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES

; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; CURRENT FILING DATE: 2001-10-31

; PRIOR APPLICATION NUMBER: US 60/244,541
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 29
; TYPE: PRT

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide

US-10-004-381-29

Query Match 49.0%; Score 102; DB 13; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.4e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 ELEQLRRLRLEHHPOGOREP 38
Db 2 ELEQLRRLRLEHHPOGOREP 20

RESULT 14
US-10-004-381-11

; Sequence 11, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:

; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES

; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; CURRENT FILING DATE: 2001-10-31

; PRIOR APPLICATION NUMBER: US 60/244,541
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 102
; TYPE: PRT

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide

US-10-004-381-11

Query Match 32.2%; Score 67; DB 13; Length 102;
Best Local Similarity 45.0%; Pred. No. 0.42;
Matches 18; Conservative 2; Mismatches 6; Indels 14; Gaps 3;

Qy 1 MDEKTTGMRGHHVEGLAGLEQLRRLRLEHHPOGOREP 38
Db 1 MDEKTTGMRGHHVEGLAGLEQLRRLRLEHHPOGOREP 28

Fri Jul 16 08:31:13 2004

RESULT 15
 US-10-156-761-13097
 ; Sequence 13097, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; NUMBER OF SEQ. ID NOS: 15109
 ; SEQ. ID NO 13097
 ; LENGTH: 739
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-13097

Query Match 32.0%; Score 66.5; DB 14; Length 739;
 Best Local Similarity 32.6%; Pred. No. 4.1;
 Matches 14; Conservative 5; Mismatches 13; Indels 11; Gaps 1;
 QY 1 MDEKTTGWRG-----GHVVEGLAGELBOLRLRLEHHP 32
 DB 341 LEDBATGMRGPEPTGMAGPRLARVWHSNRIEDLRALRLEHHP 383

Search completed: July 15, 2004, 20:45:04
 Job time : 36.7647 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:29:44 ; Search time 42.9176 Seconds

(without alignments)
250.172 Million cell updates/sec

Title: US-10-004-381-25_COPY_1_38

Perfect score: 208

Sequence: 1 MDEKTTGMRGHHVGEIAGLELQRLRLHHPGQGREP 38

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_290404:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	208	100.0	38	5	ABG67087 Streptavi
2	208	100.0	38	6	ABU57609 Streptavi
3	208	100.0	47	5	ABG67079 Streptavi
4	208	100.0	47	6	ABU57601 Synthetic
5	208	100.0	61	5	ABG67078 Streptavi
6	208	100.0	61	6	ABU57600 Synthetic
7	208	100.0	75	5	ABG67077 Streptavi
8	208	100.0	75	6	ABU57599 Synthetic
9	208	100.0	89	5	ABG67076 Streptavi
10	208	100.0	89	6	ABU57598 Synthetic
11	208	100.0	101	5	ABG67075 Streptavi
12	208	100.0	101	6	ABU57073 Streptavi
13	208	100.0	101	6	ABU57597 Streptavi
14	208	100.0	101	6	ABU57595 Streptavi
15	208	100.0	426	5	ABG67091 Maltose-b
16	208	100.0	426	6	ABU57613 Streptavi
17	208	100.0	479	5	ABG67089 Maltose-b
18	208	100.0	479	6	ABU57611 Streptavi
19	193	92.8	47	5	ABG67080 Streptavi
20	193	92.8	47	6	ABU57602 Streptavi
21	149	71.6	39	5	ABG67061 Streptavi
22	149	71.6	39	6	ABU57603 Synthetic
23	127	61.1	89	5	ABG67082 Streptavi
24	127	61.1	89	6	ABU57604 Synthetic
25	102	49.0	29	5	ABG67083 Streptavi

26	102	49.0	29	6	ABU57605 Synthetic
27	67	32.2	102	5	ABG67065 Streptavi
28	67	32.2	102	6	ABU57587 Streptavi
29	66	31.7	101	5	ABG67066 Streptavi
30	66	31.7	101	6	ABU57588 Streptavi
31	62	29.8	164	4	ABG63619 Human gas
32	60	28.8	741	4	ABG25060 Novel hum
33	59	28.4	139	6	ADA55387 Human pro
34	59	28.4	140	2	AAW40203
35	59	28.4	342	3	ABU15903
36	59	28.4	342	6	ABU14680 Protein e
37	59	28.4	538	4	ABG25106 Novel hum
38	58.5	28.1	145	5	ABP09375 Human ORF
39	58	27.9	288	6	ABR96152 Human NOV
40	58	27.9	306	4	AAZ27230 Human EXM
41	58	27.9	314	6	ABR96149 Human NOV
42	58	27.9	318	6	ABR96159 Human NOV
43	58	27.9	318	6	ABR96161 Human NOV
44	58	27.9	318	6	ABR96160 Human NOV
45	58	27.9	318	6	ABR96162 Human NOV

ALIGNMENTS

RESULT 1	ABG67087	standard; peptide; 38 AA.
XX	AC	ABG67087;
XX	DT	24-SEP-2002 (first entry)
XX	DE	Streptavidin-binding peptide and affinity tag.
XX	KW	Streptavidin-binding peptide.
OS	XX	Synthetic.
PN	XX	WO200238580-A1.
PD	XX	16-MAY-2002.
XX	PF	31-OCT-2000; 2000MO-US041717.
XX	PR	31-OCT-2000; 2000MO-US041717.
XX	PA	(GENO) GEN HOSPITAL CORP.
PI	XX	Szostrak JM, Wilson DS, Keefe AD;
DR	XX	WPI; 2002-500155/53.
PT	XX	Novel peptide with high affinity for streptavidin, is expressed as part of fusion protein to facilitate detection, quantitation and purification of desired protein.
PS	XX	Claim 13; Fig 7A; 56pp; English.
CC	XX	The invention describes a peptide (I) which binds streptavidin with a dissociation constant less than 10 mM or 23 nM, where the amino acid sequence of (I) does not contain an HPO, HPM, HPN or HOP motif, is not disulfide bonded or cyclised. A fusion protein comprising a protein of interest covalently linked to (I) is useful for purifying a desired protein from a sample by contacting the sample with streptavidin under conditions that allow complex formation between the fusion protein and the streptavidin, isolating the complex and recovering the fusion protein, and thus purifying the desired protein from the sample. The invention describes a method for producing a streptavidin-binding fusion protein. This sequence represents a streptavidin-binding peptide used in the creation of streptavidin-binding DNA-tagged peptide described in the invention

XX Synthetic peptide SB19 C-terminal deletion mutant, C4.
 XX Streptavidin-binding peptide; mRNA display; peptide library;
 KW fusion protein; mutant; mutein; SB19.
 XX Synthetic.
 OS
 PN US2002155578-A1.
 XX
 PD 24-OCT-2002.
 XX
 PF 31-OCT-2001; 2001US-00004381.
 XX
 PR 31-OCT-2000; 2000US-0244541P.
 XX
 PA (SZOS/) SZOSTAK J W.
 PA (WILS/) WILSON D S.
 PA (KEEF/) KEEFE A D.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;
 DR WPI; 2003-182639/18.
 XX
 PT Novel peptide which binds streptavidin with high affinity useful as
 PT affinity tags for purifying fusion proteins containing proteins of
 PT interest.
 XX
 PS Claim 14; Fig 5; 36pp; English.
 XX
 CC The invention relates to a peptide which binds streptavidin with a
 CC dissociation constant less than 10 micro M, where the peptide is not
 CC disulfide bonded or cyclised, or with a dissociation constant less than
 CC 10 micro M, where the amino acid sequence of the peptide does not contain
 CC an HQQ, HPW, HPN or HQP motif. The peptides are generated by the mRNA
 CC display method to create a library of peptides. Also include are a
 CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
 CC a fusion protein (and its encoding nucleic acid/vector) comprising a
 CC protein of interest covalently linked to the streptavidin binding peptide
 CC and detecting the presence of the fusion protein in a sample using
 CC streptavidin or streptavidin-containing compounds. The fusion protein
 CC (which comprises the desired protein of interest covalently linked to the
 CC streptavidin binding peptide) is useful for purifying a protein of
 CC interest from a sample, which involves expressing the fusion protein in
 CC the sample, contacting the sample with streptavidin under conditions that
 CC allow complex formation between the fusion protein and the streptavidin,
 CC isolating the complex and recovering the fusion protein, thereby
 CC purifying the protein of interest from the sample. The present sequence
 CC is a mutated version of SB19, a streptavidin binding peptide of the
 CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
 CC deletions and a truncated N-terminal peptide with point mutations
 CC
 XX
 SQ Sequence 47 AA;
 Query Match 100.0%; Score 208; DB 6; Length 47;
 Best Local Similarity 100.0%; Pred. No. 5.2e-22;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

KW Streptavidin-binding peptide; mutant; mutein.
 XX
 OS Escherichia coli.
 XX
 PN WO200238580-A1.
 XX
 PD 16-MAY-2002.
 XX
 PF 31-OCT-2000; 2000WO-US041717.
 XX
 PR 31-OCT-2000; 2000WO-US041717.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Szostak JW, Wilson DS, Keefe AD;
 DR WPI; 2002-500155/53.
 XX
 PT Novel peptide with high affinity for streptavidin, is expressed as part
 PT of fusion protein to facilitate detection, quantitation and purification
 PT of desired protein.
 XX
 PS Claim 9; Fig 5; 56pp; English.
 XX
 CC The invention describes a peptide (I) which binds streptavidin with a
 CC dissociation constant less than 10 mM or 23 nM, where the amino acid
 CC sequence of (I) does not contain an HQQ, HPW, HPN or HQP motif, is not
 CC disulfide bonded or cyclised. A fusion protein comprising a protein of
 CC interest covalently linked to (I) is useful for purifying a desired
 CC protein from a sample by contacting the sample with streptavidin under
 CC conditions that allow complex formation between the fusion protein and
 CC the streptavidin, isolating the complex and recovering the fusion
 CC protein, and thus purifying the desired protein from the sample. The
 CC invention describes a method for producing a streptavidin-binding fusion
 CC protein. This sequence represents a truncation mutant of the streptavidin
 CC -binding peptide SB-19 isolated from Escherichia coli using a method
 CC described in the specification
 CC
 XX
 SQ Sequence 61 AA;
 Query Match 100.0%; Score 208; DB 5; Length 61;
 Best Local Similarity 100.0%; Pred. No. 7e-22;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRLEHHHPOGQREP 38
 DB 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRLEHHHPOGQREP 38
 RESULT 5
 ABG67078 standard; peptide; 61 AA.
 AC
 XX
 AC ABG67078;
 XX
 DT 24-SEP-2002 (first entry)
 XX
 PR Streptavidin-binding peptide SB19 truncation mutant #4.
 XX

KW Streptavidin-binding peptide; mRNA display; peptide library;
 KW fusion protein; mutant; mutein; SB19.
 XX
 OS Synthetic.
 XX
 PN US2002155578-A1.
 XX
 PD 24-OCT-2002.
 XX
 PF 31-OCT-2001; 2001US-00004381.
 XX
 PR 31-OCT-2000; 2000US-0244541P.
 XX
 PI (SZOS/) SZOSTAK J W.
 XX

PA (WILS/) WILSON D S.
XX (KEEF/) KEEFE A D.
PI Szoestak JW, Wilson DS, Keefe AD;
XX WPI; 2003-182639/18.
XX
XX Novel peptide which binds streptavidin with high affinity useful as
PT affinity tags for purifying fusion proteins containing proteins of
PT interest.
XX
XX Claim 14; Fig 5; 36pp; English.
XX
XX The invention relates to a peptide which binds streptavidin with a
CC dissociation constant less than 10 micro M, where the peptide is not
CC disulphide bonded or cyclised, or with a dissociation constant less than
CC 10 micro M, where the amino acid sequence of the peptide does not contain
CC an HQQ, HPM, HBN or HQP motif. The peptides are generated by the mRNA
CC display method to create a library of peptides. Also include are a
CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
CC a fusion protein (and its encoding nucleic acid/vector) comprising a
CC protein of interest covalently linked to the streptavidin binding peptide
CC and detecting the presence of the fusion protein in a sample using
CC streptavidin or streptavidin-containing compounds. The fusion protein
CC (which comprises the desired protein of interest covalently linked to the
CC streptavidin binding peptide) is useful for purifying a protein of
CC interest from a sample, which involves expressing the fusion protein in
CC the sample, contacting the sample with streptavidin under conditions that
CC allow complex formation between the fusion protein and the streptavidin,
CC isolating the complex and recovering the fusion protein, thereby
CC purifying the protein of interest from the sample. The present sequence
CC is a mutated version of SB19, a streptavidin binding peptide of the
CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
CC deletions and a truncated N-terminal peptide with point mutations
XX
SQ Sequence 61 AA;
XX
Query Match 100.0%; Score 208; DB 6; Length 61;
Best Local Similarity 100.0%; Pred. No. 7e-22;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MDEKTTGWRGHHVVEGLAGELEQLRARLEHHHPQGOREP 38
DB 1 MDEKTTGWRGHHVVEGLAGELEQLRARLEHHHPQGOREP 38
XX
RESULT 7
ABG67077 standard; peptide; 75 AA.
XX
AC ABG67077;
XX
DT 24-SEP-2002 (first entry)
XX
DE Streptavidin-binding peptide SB19 truncation mutant #3.
XX
KM Streptavidin-binding peptide; mutant; mutein.
XX
OS Escherichia coli.
XX
PN WO200238580-A1.
XX
PD 16-MAY-2002.
XX
PF 31-OCT-2000; 2000WO-US041717.
XX
PR 31-OCT-2000; 2000WO-US041717.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Szoestak JW, Wilson DS, Keefe AD;
XX
DR WPI; 2002-500155/53.

XX
XX Novel peptide with high affinity for streptavidin, is expressed as part
PT of fusion protein to facilitate detection, quantitation and purification
PT of desired protein.
XX
XX Claim 9; Fig 5; 56pp; English.
XX
XX The invention describes a peptide (I) which binds streptavidin with a
CC dissociation constant less than 10 nM or 23 nM, where the amino acid
CC sequence of (I) does not contain an HQQ, HPM, HBN or HQP motif, is not
CC disulphide bonded or cyclised. A fusion protein comprising a protein of
CC interest covalently linked to (I) is useful for purifying a desired
CC protein from a sample by contacting the sample with streptavidin under
CC conditions that allow complex formation between the fusion protein and
CC the streptavidin, isolating the complex and recovering the fusion
CC protein, and thus purifying the desired protein from the sample. The
CC invention describes a method for producing a streptavidin-binding fusion
CC protein. This sequence represents a truncation mutant of the streptavidin
CC -binding peptide SB-19 isolated from Escherichia coli using a method
CC described in the specification
XX
SQ Sequence 75 AA;
XX
Query Match 100.0%; Score 208; DB 5; Length 75;
Best Local Similarity 100.0%; Pred. No. 8.8e-22;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MDEKTTGWRGHHVVEGLAGELEQLRARLEHHHPQGOREP 38
DB 1 MDEKTTGWRGHHVVEGLAGELEQLRARLEHHHPQGOREP 38
XX
RESULT 8
ABU57599 standard; peptide; 75 AA.
XX
AC ABU57599;
XX
DT 09-APR-2003 (first entry)
XX
DE Synthetic peptide SB19 C-terminal deletion mutant, C2.
XX
KM Streptavidin-binding peptide; mRNA display; peptide library;
XX fusion protein; mutant; mutein; SB19.
XX
OS Synthetic.
XX
PN US2002155578-A1.
XX
PD 24-OCT-2002.
XX
PF 31-OCT-2001; 2001US-00004381.
XX
PR 31-OCT-2000; 2000US-0244541P.
XX
PA (SZOS/) SZOSTAK J W.
XX (WILS/) WILSON D S.
XX (KEEF/) KEEFE A D.
XX
XX Szoestak JW, Wilson DS, Keefe AD;
XX
XX WPI; 2003-182639/18.
XX
XX Novel peptide which binds streptavidin with high affinity useful as
PT affinity tags for purifying fusion proteins containing proteins of
PT interest.
XX
XX Claim 14; Fig 5; 36pp; English.
XX
XX The invention relates to a peptide which binds streptavidin with a
CC dissociation constant less than 10 micro M, where the peptide is not
CC disulphide bonded or cyclised, or with a dissociation constant less than
CC 10 micro M, where the amino acid sequence of the peptide does not contain

CC an HPO, HPM, HPN or HOP motif. The peptides are generated by the mRNA
CC display method to create a library of peptides. Also include are a
CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
CC a fusion protein (and its encoding nucleic acid/vector) comprising a
CC protein of interest covalently linked to the streptavidin binding peptide
CC and detecting the presence of the fusion protein in a sample using
CC streptavidin or streptavidin-containing compounds. The fusion protein
CC (which comprises the desired protein of interest covalently linked to the
CC streptavidin binding peptide) is useful for purifying a protein of
CC interest from a sample, which involves expressing the fusion protein in
CC the sample, contacting the sample with streptavidin under conditions that
CC allow complex formation between the fusion protein and the streptavidin,
CC isolating the complex and recovering the fusion protein, thereby
CC purifying the protein of interest from the sample. The present sequence
CC is a mutated version of SB19, a streptavidin binding peptide of the
CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
CC deletions and a truncated N-terminal peptide with point mutations
SQ Sequence 75 AA;

Query Match 100.0%; Score 208; DB 6; Length 75;
Best Local Similarity 100.0%; Pred. No. 8.8e-22;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGQREP 38
Db 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGQREP 38

RESULT 9
ABG67076 standard; peptide; 89 AA.

AC ABG67076;
DT 24-SEP-2002 (first entry)

DE Streptavidin-binding peptide SB19 truncation mutant #2.

KM Streptavidin-binding peptide; mutant; mutein.

OS Escherichia coli.

PN MO200238580-A1.

PD 16-MAY-2002.

PF 31-OCT-2000; 2000MO-US041717.

PR 31-OCT-2000; 2000MO-US041717.

PA (GEHO) GEN HOSPITAL CORP.

PI Szostak JW, Wilson DS, Keefe AD;

DR WPI; 2002-500155/53.

PT Novel peptide with high affinity for streptavidin, is expressed as part
PT of fusion protein to facilitate detection, quantitation and purification
PT of desired protein.

PS Claim 9; Fig 5; 56pp; English.

CC The invention describes a peptide (I) which binds streptavidin with a
CC dissociation constant less than 10 mM or 23 nM, where the amino acid
CC sequence of (I) does not contain an HPO, HPM, HPN or HOP motif, is not
CC disulfide bonded or cyclized. A fusion protein comprising a protein of
CC interest covalently linked to (I) is useful for purifying a desired
CC protein from a sample by contacting the sample with streptavidin under
CC conditions that allow complex formation between the fusion protein and
CC the streptavidin, isolating the complex and recovering the fusion
CC protein, and thus purifying the desired protein from the sample. The
CC invention describes a method for producing a streptavidin-binding fusion

CC protein. This sequence represents a truncation mutant of the streptavidin
CC -binding peptide SB-19 isolated from Escherichia coli using a method
CC described in the specification

SQ Sequence 89 AA;

Query Match 100.0%; Score 208; DB 5; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.1e-21;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGQREP 38
Db 1 MDEKTTGMRGHHVVEGLAGLEQLRLRLHHHPQGQREP 38

RESULT 10
ABU57598 standard; peptide; 89 AA.

AC ABU57598;

DT 09-APR-2003 (first entry)

DE Synthetic peptide SB19 C-terminal deletion mutant, C1.

KM Streptavidin-binding peptide; mRNA display; peptide library;
KM fusion protein; mutant; mutein; SB19.

OS Synthetic.

PN US200215578-A1.

PD 24-OCT-2002.

PF 31-OCT-2001; 2001US-00004381.

PR 31-OCT-2000; 2000US-0244541P.

PA (SZOS/) SZOSTAK J W.

PA (WILS/) WILSON D S.

PA (KEEF/) KEEFE A D.

PI Szostak JW, Wilson DS, Keefe AD;

DR WPI; 2003-182639/18.

PT Novel peptide which binds streptavidin with high affinity useful as
PT affinity tags for purifying fusion proteins containing proteins of
PT interest.

PS Claim 14; Fig 5; 36pp; English.

CC The invention relates to a peptide which binds streptavidin with a
CC dissociation constant less than 10 micro M, where the peptide is not
CC disulfide bonded or cyclized, or with a dissociation constant less than
CC 10 micro M, where the amino acid sequence of the peptide does not contain
CC an HPO, HPM, HPN or HOP motif. The peptides are generated by the mRNA
CC display method to create a library of peptides. Also include are a
CC nucleic acid encoding the peptide, a vector comprising the nucleic acid,
CC a fusion protein (and its encoding nucleic acid/vector) comprising a
CC protein of interest covalently linked to the streptavidin binding peptide
CC and detecting the presence of the fusion protein in a sample using
CC streptavidin or streptavidin-containing compounds. The fusion protein
CC (which comprises the desired protein of interest covalently linked to the
CC streptavidin binding peptide) is useful for purifying a protein of
CC interest from a sample, which involves expressing the fusion protein in
CC the sample, contacting the sample with streptavidin under conditions that
CC allow complex formation between the fusion protein and the streptavidin,
CC isolating the complex and recovering the fusion protein, thereby
CC purifying the protein of interest from the sample. The present sequence
CC is a mutated version of SB19, a streptavidin binding peptide of the
CC invention. The mutants comprise 3 N-terminal deletions, 3 C-terminal
CC deletions and a truncated N-terminal peptide with point mutations

XX Sequence 89 AA;
SQ

Query Match 100.0%; Score 208; DB 6; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.1e-21;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDEKTTGWRGHHVVEGLAGLEQLRRLRHHHPQGQREP 38
Db 1 MDEKTTGWRGHHVVEGLAGLEQLRRLRHHHPQGQREP 38

RESULT 11

ABG67075
ID ABG67075 standard; peptide; 101 AA.

XX ABG67075;

XX 24-SEP-2002 (first entry)

XX Streptavidin-binding peptide SB19 truncation mutant #1.

XX Streptavidin-binding peptide; mutant; mutein.

XX Escherichia coli.

OS Synthetic.

XX WO200238580-A1.

XX 16-MAY-2002.

XX 31-OCT-2000; 2000WO-US041717.

XX 31-OCT-2000; 2000WO-US041717.

XX (GENO) GEN HOSPITAL CORP.

XX Szostak JW, Wilson DS, Keefe AD;

XX MPI; 2002-500155/53.

PT Novel peptide with high affinity for streptavidin, is expressed as part
PT of fusion protein to facilitate detection, quantitation and purification
PT of desired protein.

PS Claim 9; Fig 5; 56pp; English.

CC The invention describes a peptide (I) which binds streptavidin with a
CC dissociation constant less than 10 mM or 23 nM, where the amino acid
CC sequence of (I) does not contain an HPQ, HPW, HPN or HQP motif, is not
CC disulfide bonded or cyclised. A fusion protein comprising a protein of
CC interest covalently linked to (I) is useful for purifying a desired
CC protein from a sample by contacting the sample with streptavidin under
CC conditions that allow complex formation between the fusion protein and
CC the streptavidin, isolating the complex and recovering the fusion
CC protein, and thus purifying the desired protein from the sample. The
CC invention describes a method for producing a streptavidin-binding fusion
CC protein. This sequence represents a truncation mutant of the streptavidin
CC -binding peptide SB-19 isolated from Escherichia coli using a method
CC described in the specification

XX SQ Sequence 101 AA;

Query Match 100.0%; Score 208; DB 5; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.2e-21;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDEKTTGWRGHHVVEGLAGLEQLRRLRHHHPQGQREP 38
Db 1 MDEKTTGWRGHHVVEGLAGLEQLRRLRHHHPQGQREP 38

RESULT 12

ABG67073
ID ABG67073 standard; peptide; 101 AA.

XX ABG67073;

XX 24-SEP-2002 (first entry)

XX Streptavidin-binding peptide SB19.

XX Streptavidin-binding peptide.

XX Escherichia coli.

OS WO200238580-A1.

XX 16-MAY-2002.

XX 31-OCT-2000; 2000WO-US041717.

XX 31-OCT-2000; 2000WO-US041717.

XX (GENO) GEN HOSPITAL CORP.

XX Szostak JW, Wilson DS, Keefe AD;

XX MPI; 2002-500155/53.

PT Novel peptide with high affinity for streptavidin, is expressed as part
PT of fusion protein to facilitate detection, quantitation and purification
PT of desired protein.

PS Claim 9; Fig 3; 56pp; English.

CC The invention describes a peptide (I) which binds streptavidin with a
CC dissociation constant less than 10 mM or 23 nM, where the amino acid
CC sequence of (I) does not contain an HPQ, HPW, HPN or HQP motif, is not
CC disulfide bonded or cyclised. A fusion protein comprising a protein of
CC interest covalently linked to (I) is useful for purifying a desired
CC protein from a sample by contacting the sample with streptavidin under
CC conditions that allow complex formation between the fusion protein and
CC the streptavidin, isolating the complex and recovering the fusion
CC protein, and thus purifying the desired protein from the sample. The
CC invention describes a method for producing a streptavidin-binding fusion
CC protein. This sequence represents a streptavidin-binding peptide isolated
CC from Escherichia coli using a method described in the specification

XX SQ Sequence 101 AA;

Query Match 100.0%; Score 208; DB 5; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.2e-21;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDEKTTGWRGHHVVEGLAGLEQLRRLRHHHPQGQREP 38
Db 1 MDEKTTGWRGHHVVEGLAGLEQLRRLRHHHPQGQREP 38

RESULT 13

ABU57597
ID ABU57597 standard; peptide; 101 AA.

XX ABU57597;

XX 09-APR-2003 (first entry)

XX Streptavidin-binding synthetic peptide SB19 full length peptide.

XX Streptavidin-binding peptide; mRNA display; peptide library;

XX fusion protein.

OS Synthetic.

XX US2002155578-A1.

DR WPI; 2002-500155/53.
DR N-PSDB; ABK95732.
XX

PT Novel peptide with high affinity for streptavidin, is expressed as part
PT of fusion protein to facilitate detection, quantitation and purification
PT of desired protein.
XX

PS Disclosure; Fig 9B; 56pp; English.
XX

CC The invention describes a peptide (I) which binds streptavidin with a
CC dissociation constant less than 10 mM or 23 nM, where the amino acid
CC sequence of (I) does not contain an HPQ, HPN, HPN or HQP motif, is not
CC disulfide bonded or cyclised. A fusion protein comprising a protein of
CC interest covalently linked to (I) is useful for purifying a desired
CC protein from a sample by contacting the sample with streptavidin under
CC conditions that allow complex formation between the fusion protein and
CC the streptavidin, isolating the complex and recovering the fusion
CC protein, and thus purifying the desired protein from the sample. The
CC invention describes a method for producing a streptavidin-binding fusion
CC protein. This sequence represents a fusion protein comprising maltose-
CC binding protein, a streptavidin-binding peptide, and a hexahistidine tag
XX

SO Sequence 426 AA;

Query Match 100.0%; Score 208; DB 5; Length 426;
Best Local Similarity 100.0%; Pred. No. 6.2e-21;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MDEKTTGARGSHVVEGLAGLEQLRLRHHHPQGQREP 38
|||
DB 378 MDEKTTGARGSHVVEGLAGLEQLRLRHHHPQGQREP 415

Search completed: July 15, 2004, 20:34:02
Job time : 42.9176 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:31:20 ; Search time 12.7176 Seconds
(without alignments)
355.490 Million cell updates/sec

Title: US-10-004-381-25

Sequence: 1 MDEKTTGWRGSHVEGLAGE.....LEHHPQGGREPMMSGGCKLG 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

```
Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	73.5	28.5	554	2	E87375	conserved hypothe
2	65.5	25.4	528	2	C87663	conserved hypothe
3	61	23.6	429	2	F84015	maltose/maltodextr
4	59.5	23.1	342	2	E64905	probable sugar tra
5	59.5	23.1	342	2	A90894	probable transport
6	59.5	23.1	342	2	H85723	probable transport
7	59	22.9	1974	2	T30010	hypothetical prote
8	58.5	22.7	302	2	B82189	ROK family protein
9	57	22.1	474	1	E24723	trpc protein - Cor
10	57	22.1	496	2	H70668	probable polycetid
11	57	22.1	663	2	D83545	probable helicase
12	56.5	21.9	215	2	A11074	phosphoglycerate p
13	56.5	21.9	1560	2	I54361	SMCX protein - hum
14	56	21.7	261	2	S43218	hypothetical prote
15	56	21.7	338	2	T36025	conserved hypothet
16	55.5	21.5	188	2	E87711	heat shock protein
17	55.5	21.5	288	2	E72671	hypothetical prote
18	55.5	21.5	392	2	F84650	probable protein p
19	55	21.3	324	2	AE1364	protein gp49 (Bact
20	55	21.3	377	2	A95265	probable dehydratr
21	54	20.9	147	2	F83505	probable peptide d
22	54	20.9	321	2	H87683	transcription regul
23	54	20.9	1156	2	E69444	chromosome segrega
24	53.5	20.7	134	2	AH3177	conserved hypothet
25	53.5	20.7	352	2	H71294	probable o-sialogl
26	53.5	20.7	404	2	S68409	potassium channel
27	53.5	20.7	478	2	C86274	hypothetical prote
28	53	20.5	210	2	B70559	hypothetical prote
29	53	20.5	224	2	D83570	probable nucleotid

30	53	20.5	261	2	C75568	hypotheical prot
31	53	20.5	348	1	S75456	protein plex - syn
32	53	20.5	358	2	C72645	probable 3-dehydro
33	53	20.5	528	2	C85056	probable DNA-bind
34	53	20.5	589	2	S06954	intermediate filam
35	53	20.5	947	2	B86362	hypotheical prote
36	52.5	20.3	184	2	A13507	heat shock protein
37	52.5	20.3	744	2	A82822	NADH-ubiquinone ox
38	52	20.2	151	2	J00810	transcription repr
39	52	20.2	356	2	T22478	hypocheical prote
40	52	20.2	376	2	T34734	hypotheical prote
41	52	20.2	378	2	A25399	homeotic protein A
42	52	20.2	450	2	T35496	probable 2-dehydro
43	52	20.2	494	2	I58658	neurofilament -66
44	52	20.2	504	2	I53668	alpha-internexin -
45	52	20.2	505	2	A41023	alpha-internexin -

ALIGNMENTS

RESULT 1
E87375
conserved hypothetical protein CCI017 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87375
R:Nierman, W.C.; Felblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Bisen, J.; Heidelberg, J.
B.; Ijamb, M.T.; Deboy, R.T.; Dodson, R.J.; Duxkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; M01D:21173698; EMD:11259647
A:Accession: E87375
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-554 <STO>
A:CROSS-references: GB:AE005673; NID:g13422309; PIDN:AAK3001.1, GSPDB:GN00148
;Genetics:
;Gene: CCI017

```

Query Match          28:5#; Score 73.5; DB 2; Length 554;
Best Local Similarity 35.4#; Pred. No. 0.27;
Matches 17; Conservative 6; Mismatches 20; Indels 5; Gaps 1;

Ox      5 TTGARGSHVVE---GLAGELPOLRLARLHHHPQGRREPMSSGCKKLG 47
       :|::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      45 STEMSAGYVDVNVYTFPGYVGEINPLRCRLPLTLTVGRHAPKTENACEBG 92

RESULT 2
C87663
Conserved hypothetical protein CCJ341 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence-revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: C87663
R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, N.J.; Rmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Frazer, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: AB7249; MUID:21173698; PMID:11259647
A:Accession: C87663
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-528 <STO>
C:Cross-references: GB:AE005673; NID:GI3425041; PIDN:AAK25303.1; GSPDB:GN00148
C:Genetics:
A:Gene: CCJ341

Query Match          25:4#; Score 65.5; DB 2; Length 528;
Best Local Similarity 37.8#; Pred. No. 2.6;
Matches 17; Conservative 5; Mismatches 18; Indels 5; Gaps 1;

```


A:Molecule type: DNA
A:Residues: 1-1974 <DUZ>
A:Cross-references: EMBL:U50309; PIDD:AA037057.1; GSPDB:GN00023; CESP:F58G4.1
A:Experimental source: strain Bristol N2; clone F58G4
C:Genetics:
A:Gene: CESP:F58G4.1
A:Map position: 5
A:Insertions: 18/3; 111/3; 164/1; 229/1; 264/1; 378/1; 440/2; 525/3; 1177/2; 1633/3; 1863/3
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:84-776/Domain: myosin motor domain homology <MMO>

Query Match 22.9%; Score 59; DB 2; Length 1974;
Best Local Similarity 35.7%; Pred. No. 67;
Matches 15; Conservative 7; Mismatches 18; Indels 2; Gaps 1;

QY 8 WRCGHVVEGL--AGELEQLARLEHHPOGOREPMMSGCCTLG 47
DB 1381 WRKFEQEGVSRABELETTRKLTHTVQEQOLENANOKIG 1422

RESULT 8

B82189
ROK family protein VC1532 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82189
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.D.;
Churchson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bais, S.; Qin, H.; Dragol, I.; Sellers, F.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: B82189
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-302 <HEI>
A:Cross-references: GB:AE004231; GB:AE003852; NID:99556027; PIDD:AA94686.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1532
A:Map position: 1
C:Superfamily: conserved hypothetical protein H10182; glucose kinase homology

Query Match 22.7%; Score 58.5; DB 2; Length 302;
Best Local Similarity 32.7%; Pred. No. 11;
Matches 16; Conservative 10; Mismatches 12; Indels 11; Gaps 3;

QY 6 TGRGCHVVEG-----LAGLEQLARLE--HHPOGOREPMMSGC 44
DB 132 TGRGCHVVEGKVFSGRNHAGETIGMRPLIDMFF-LGRKAPLJAGCC 179

RESULT 9

E24723
trpC protein - Corynebacterium glutamicum
C:Species: Corynebacterium glutamicum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: E24723
R:Matsumi, K.; Sano, K.; Ohtsubo, E.
Nucleic Acids Res. 14, 10113-10114, 1986
A:Title: Complete nucleotide and deduced amino acid sequences of the Brevibacterium lact
A:Reference number: A93606; MUID:87117512; PMID:3808947
A:Contents: B. lactofermentum
A:Accession: E24723
A:Molecule type: DNA
A:Residues: 1-474 <NAT>
A:Cross-references: GB:X04960; NID:g39591; PIDD:CAA28626.1; PID:g39596
C:Genetics:
A:Gene: trpC
C:Superfamily: trpC-trpF bifunctional enzyme; trpC homology; trpF homology
F:9-258/Domain: trpC homology <TRC>
F:264-460/Domain: trpF homology <TRF>

Query Match 22.1%; Score 57; DB 1; Length 474;
Best Local Similarity 55.6%; Pred. No. 27;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 13 VVEGLAGELEQLARLEH 30
DB 13 IVEGRGHLEETRIARAH 30

RESULT 10

H70668
probable polyketide synthase - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70668; S73074
R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
J.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70668
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-496 <COU>
A:Cross-references: GB:Z83859; GB:AL123456; NID:93261675; PIDD:CA06102.1; PID:g1781166
A:Experimental source: strain H37RV
R:Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, September 1994
A:Description: Mycobacterium tuberculosis cosmid tbcz.
A:Reference number: S73053
A:Accession: S73074
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-273; 275-496 <SNI>
A:Cross-references: EMBL:U00024; NID:g560506; PIDD:AAA50929.1; PID:g560508
C:Genetics:
A:Gene: pks15
A:Start codon: GTG
C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I; 3-oxoacyl-[acyl-carrier-prot
F:86-467/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>

Query Match 22.1%; Score 57; DB 2; Length 496;
Best Local Similarity 36.1%; Pred. No. 28;
Matches 13; Conservative 9; Mismatches 12; Indels 2; Gaps 1;

QY 12 HVEGLAGELEQLARLEHHPOGOREP--MMSGCK 45
DB 22 HYLKRVAVELDETRRLREYEQRAPEVAVVGICR 57

RESULT 11

D83545
probable helicase PA0799 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83545
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, L.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lam,
., Loy, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: D83545
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-663 <STO>
A:Cross-references: GB:AE004515; GB:AE004091; NID:g9946687; PIDD:AA04188.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0799

Fri Jul 16 08:31:15 2004

us-10-004-381-25.rpr

Page 5

Db 226 KYG 228

Search completed: July 15, 2004, 20:36:20
Job time : 13.7176 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:35:51 ; Search time 44.2353 Seconds
(without alignments)
332.096 Million cell updates/sec

Title: US-10-004-381-25

Perfect score: 258
Sequence: 1 MDEKTTGWRGHHVVEGLAGE.....LEHHPGQGRPMWGGCKLG 47

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubppa/US09C_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubppa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubppa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	258	100.0	47	13	US-10-004-381-25 Sequence 25, Appl
2	243	94.2	47	13	US-10-004-381-26 Sequence 26, Appl
3	241	93.4	61	13	US-10-004-381-24 Sequence 24, Appl
4	234	90.7	75	13	US-10-004-381-23 Sequence 23, Appl
5	227	88.0	89	13	US-10-004-381-22 Sequence 22, Appl
6	214	82.9	101	13	US-10-004-381-21 Sequence 21, Appl
7	211	81.8	101	13	US-10-004-381-19 Sequence 19, Appl
8	208	80.6	38	13	US-10-004-381-35 Sequence 35, Appl
9	208	80.6	426	13	US-10-004-381-41 Sequence 41, Appl
10	208	80.6	479	13	US-10-004-381-38 Sequence 38, Appl
11	199	77.1	39	13	US-10-004-381-27 Sequence 27, Appl
12	152	58.9	29	13	US-10-004-381-29 Sequence 29, Appl
13	133	51.6	89	13	US-10-004-381-28 Sequence 28, Appl
14	94	36.4	44	13	US-10-004-381-34 Sequence 34, Appl
15	70	27.1	102	13	US-10-004-381-11 Sequence 11, Appl

16	68	26.4	101	13	US-10-004-381-12	Sequence 12, Appl
17	66.5	25.8	739	14	US-10-156-761-13097	Sequence 13097, A
18	65	25.2	587	12	US-10-425-114-73019	Sequence 73019, A
19	62.5	24.2	175	16	US-10-437-963-147749	Sequence 147749, A
20	62	24.0	114	12	US-10-425-114-61050	Sequence 61050, A
21	62	24.0	527	12	US-10-425-114-70961	Sequence 70961, A
22	62	24.0	529	12	US-10-425-114-77475	Sequence 37475, A
23	62	24.0	649	12	US-10-425-114-70144	Sequence 70144, A
24	61	23.6	1181	14	US-10-156-761-12448	Sequence 12448, A
25	59.5	23.1	84	11	US-09-864-408A-6610	Sequence 6610, Ap
26	59.5	23.1	342	9	US-09-912-020-260	Sequence 260, App
27	59.5	23.1	342	12	US-10-282-122A-42604	Sequence 42604, A
28	59	22.9	139	15	US-10-094-749-2955	Sequence 2955, Ap
29	59	22.9	140	9	US-09-902-180-6	Sequence 6, Appl
30	59	22.9	147	9	US-09-864-761-39307	Sequence 39307, A
31	59	22.9	149	16	US-10-437-963-170766	Sequence 170766, A
32	59	22.9	160	16	US-10-437-963-161643	Sequence 161643, A
33	59	22.9	249	12	US-10-112-944-859	Sequence 859, App
34	59	22.9	249	12	US-10-112-944-860	Sequence 860, App
35	59	22.9	249	12	US-10-112-944-861	Sequence 861, App
36	59	22.9	249	12	US-10-112-944-862	Sequence 862, App
37	59	22.9	869	16	US-10-437-963-156107	Sequence 156107, A
38	59	22.9	1974	15	US-10-369-493-6395	Sequence 6395, Ap
39	58.5	22.7	1336	16	US-10-437-963-115864	Sequence 115864, A
40	58	22.5	115	12	US-10-282-122A-45153	Sequence 45153, A
41	58	22.5	127	16	US-10-437-963-178804	Sequence 178804, A
42	58	22.5	205	12	US-10-425-114-39504	Sequence 39504, A
43	58	22.5	239	12	US-10-425-114-39359	Sequence 39359, A
44	58	22.5	276	12	US-10-425-114-38383	Sequence 38383, A
45	58	22.5	288	12	US-10-112-944-410	Sequence 410, App

ALIGNMENTS

RESULT 1
US-10-004-381-25
; Sequence 25, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-25

Query Match 100.0%; Score 258; DB 13; Length 47;
Best Local Similarity 100.0%; Pred. No. 7e-26;
Matches 47; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEKTTGWRGHHVVEGLAEQLRLARLHHHFGQGRPMWGGCKLG 47
DB 1 MDEKTTGWRGHHVVEGLAEQLRLARLHHHFGQGRPMWGGCKLG 47

RESULT 2
US-10-004-381-26
; Sequence 26, Application US/10004381
; Publication No. US20020155578A1

```
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ CURRENT FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ PRIOR FILING DATE: 2000-10-31
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 26
/ LENGTH: 47
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-26
```

```
Query Match          94.2%; Score 243; DB 13; Length 47;
Best Local Similarity 95.7%; Pred. No. 6e-24;
Matches 45; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPQGRPPMMSGGCKLG 47
DB 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPQGRPPMMSGGCKLG 47
```

```
RESULT 3
US-10-004-381-24
/ Sequence 24, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ CURRENT FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ PRIOR FILING DATE: 2000-10-31
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 24
/ LENGTH: 61
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-24
```

```
Query Match          93.4%; Score 241; DB 13; Length 61;
Best Local Similarity 77.0%; Pred. No. 1.5e-23;
Matches 47; Conservative 0; Mismatches 0; Indels 14; Gaps 1;
```

```
QY 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPQGRPPMMSGGCKLG 46
DB 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPQGRPPMMSGGCKLG 60
```

```
QY 47 G 47
DB 61 G 61
```

```
RESULT 4
US-10-004-381-23
/ Sequence 23, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
```

```
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ CURRENT FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ PRIOR FILING DATE: 2000-10-31
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 23
/ LENGTH: 75
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-23
```

```
Query Match          90.7%; Score 234; DB 13; Length 75;
Best Local Similarity 62.7%; Pred. No. 1.5e-22;
Matches 47; Conservative 0; Mismatches 0; Indels 28; Gaps 1;
```

```
QY 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPQGRPPMMSGGCKLG 38
DB 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPQGRPPMMSGGCKLG 60
```

```
QY 39 -----MMSGGCKLG 47
DB 61 LLDPEVKLLTDWPKFKKIVSMMSGGCKLG 75
```

```
RESULT 5
US-10-004-381-22
/ Sequence 22, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ CURRENT FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ PRIOR FILING DATE: 2000-10-31
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 22
/ LENGTH: 89
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-22
```

```
Query Match          88.0%; Score 227; DB 13; Length 89;
Best Local Similarity 52.8%; Pred. No. 1.4e-21;
Matches 47; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
```

```
QY 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPQGRPPMMSGGCKLG 38
DB 1 MDEKTTGMRGSHVVEGLAGLEQLRRLRHHHPQGRPPMMSGGCKLG 60
```

```
QY 39 -----MMSGGCKLG 47
DB 61 LLDPEVKLLTDWPKFKKIVSMMSGGCKLG 89
```

```
RESULT 6
US-10-004-381-21
```



```

: Sequence 21 Application US/10004381
: Publication No. US20020155578A1
:
: GENERAL INFORMATION:
: APPLICANT: SZOSTAK, JACK W.
: APPLICANT: WILSON, DAVID S.
: APPLICANT: KEEFE, ANTHONY D.
: TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
: TITLE OF INVENTION: THEREOF
: FILE REFERENCE: 00786/388002
: CURRENT APPLICATION NUMBER: US/10/004,381
: CURRENT FILING DATE: 2001-10-31
: PRIOR APPLICATION NUMBER: US 60/244,541
: PRIOR FILING DATE: 2000-10-31
: NUMBER OF SEQ ID NOS: 41
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 21
:
: LENGTH: 101
: TYPE: PRT
: ORGNISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: selected peptide
:
: US-10-004-381-21

```

Query Match	82.9%;	Score 214;	DB 13;	Length 101;
Best Local Similarity	44.6%;	Pred. No. 7.8e-20;		
Matches 45;	Conservative 2;	Mismatches 0;	Indels 54;	Gaps 1

```
QY      1 MDEKTTGWRGGVIVEGLAGELBQLRARLHHNPGQRREPM----- 40
        |||||
Db       1 MDEKTGMWGRGVVEGLAGELBQLRARLEHHPQGRREP LVQEVEDVDGLVQDLHGVAAG 60
```

QY 41 -----SGGCKLG 47
| | | | |
Db 61 LLDPEVKLTDMFKFKPNVSKDCKMTFYLEMYDWSGGCKLG 101

```

RESULT 7
US-10-004-381-19
; Sequence 19, Application US/10004381
; Publication No. US20020155578A1
; GENERAL INFORMATION:
; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEPE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIRIN-BINDING PEPTIDES AND USES
; TITLE OF INVENTION: THERBOF
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-19

```

Query Match	81.8%	Score 211;	DB 13;	Length 101;
Best Local Similarity	95.0%	Pred. No. 1.9e-19;		
Matches 38;	Conservative 2;	Mismatches 0;	Indels 0;	Gaps 0

Oy

1 MDEKTTGWRGGHVVEGLAGLEEQRLARLEHHHPQOGOREPMM 40

D8

1 MDEKTTGWRGGHVVEGLAGLEEQRLARLEHHHPQOGOREPLV 40

RESULT 8
US-10-004-381-35
; Sequence 35, Application US/10004381

```

Publication No. US20020155578A1
GENERAL INFORMATION:
APPLICANT: SZOSTAK, JACK W.
APPLICANT: WILSON, DAVID S.
APPLICANT: KEEFE, ANTHONY D.
TITLE OF INVENTION: STEPPED/AVIDIN-BINDING PEPTIDES AND USES
TITLE OF INVENTION: THERBOF
FILE REFERENCE: 00786/388002
CURRENT APPLICATION NUMBER: US/10/004,381
CURRENT FILING DATE: 2001-10-31
PRIOR APPLICATION NUMBER: US 60/244,541
PRIOR FILING DATE: 2000-10-31
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 35
LENGTH: 38
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: selected peptide
US-10-004-381-35

```

Query Match	80.6%	Score 208;	DB 13;	Length 38;
Best Local Similarity	100.0%	Pred. No. 1.6e-19;		
Matches 38; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

[illegible]

```

RESULT 9
US-10-004-381-41
Sequence 41, Application US/10004381
Publication No. US20020155578A1
GENERAL INFORMATION:
APPLICANT: SZOSTAK, JACK W.
APPLICANT: WILSON, DAVID S.
APPLICANT: KEEFE, ANTHONY D.
TITLE OF INVENTION: STEPEPAVIDIN-BINDING PEPTIDES AND USES
TITLE OF INVENTION: THERSOF
FILE REFERENCE: 00786/389002
CURRENT APPLICATION NUMBER: US/10/004,381
CURRENT FILING DATE: 2001-10-31
PRIOR APPLICATION NUMBER: US 60/244,541
PRIOR FILING DATE: 2000-10-31
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 41
LENGTH: 426
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: designed protein
US-10-004-381-41

```

Query Match	80.6%	Score 208;	DB 13;	length 426;
Best Local Similarity	100.0%	Pred. No. 2.3e-18;		
Matches 38; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

Oy 1 MDEKTGWRGSHVVEGLAGELEQLRLARLEHHHPQGRREP 38
Db 378 MDEKTTGWRCGHVVEGLAGELQLRLARLEHHHPQGRREP 415

RESULT 10
US-10-004-381-38

APPLICANT: SZOSTAK, JACK W.
APPLICANT: WILSON, DAVID S.
APPLICANT: KEEFE, ANTHONY D

```
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE OF INVENTION: THEREOF
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ PRIOR FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 38
/ LENGTH: 479
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Designed protein
US-10-004-381-38

Query Match          80.6%; Score 208; DB 13; Length 479;
Best Local Similarity 100.0%; Pred. No. 2.6e-18;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MDEKTTGMRGSHVVEGLAELEQLRLRLEHHPOGQREP 38
DB      378 MDEKTTGMRGSHVVEGLAELEQLRLRLEHHPOGQREP 415

RESULT 11
US-10-004-381-27
/ Sequence 27, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ PRIOR FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 27
/ LENGTH: 39
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-27

Query Match          77.1%; Score 199; DB 13; Length 39;
Best Local Similarity 100.0%; Pred. No. 2.4e-18;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      11 GHVVEGLAELEQLRLRLEHHPOGQREPMMSGGCKLG 47
DB      3 GHVVEGLAELEQLRLRLEHHPOGQREPMMSGGCKLG 39

RESULT 12
US-10-004-381-29
/ Sequence 29, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ PRIOR FILING DATE: 2001-10-31
```

```
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ PRIOR FILING DATE: 2000-10-31
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 29
/ LENGTH: 29
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-29

Query Match          58.9%; Score 152; DB 13; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.1e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      20 ELEQLRLRLEHHPOGQREPMMSGGCKLG 47
DB      2 ELEQLRLRLEHHPOGQREPMMSGGCKLG 29

RESULT 13
US-10-004-381-28
/ Sequence 28, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ PRIOR FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 28
/ LENGTH: 89
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: selected peptide
US-10-004-381-28

Query Match          51.6%; Score 133; DB 13; Length 89;
Best Local Similarity 35.6%; Pred. No. 2e-09;
Matches 31; Conservative 2; Mismatches 0; Indels 54; Gaps 1;

QY      15 EGLAELEQLRLRLEHHPOGQREPMM----- 40
DB      3 EGLAELEQLRLRLEHHPOGQREPLYVEDVDEGLVDLHGVAGLDPVEKLLTDMFK 62

QY      41 -----SGGCKLG 47
DB      63 KFKVSKDCKMTFYLEMVDMSGGCKLG 89

RESULT 14
US-10-004-381-34
/ Sequence 34, Application US/10004381
/ Publication No. US20020155578A1
/ GENERAL INFORMATION:
/ APPLICANT: SZOSTAK, JACK W.
/ APPLICANT: WILSON, DAVID S.
/ APPLICANT: KEEFE, ANTHONY D.
/ TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
/ FILE REFERENCE: 00786/388002
/ CURRENT APPLICATION NUMBER: US/10/004,381
/ PRIOR FILING DATE: 2001-10-31
/ PRIOR APPLICATION NUMBER: US 60/244,541
```

; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: designed peptide
US-10-004-381-34

Query Match 36.4%; Score 94; DB 13; Length 44;
Best Local Similarity 58.3%; Pred. No. 0.0001;
Matches 21; Conservative 3; Mismatches 4; Indels 8; Gaps 2;

QY 18 AGELEQLRAR-----LEHHPOGOREPMMGGCKTG 47
DB 8 AGFVDQADARLVQAGALQHHPOGDR--MMSGCKTG 41

RESULT 15
US-10-004-381-11
; Sequence 11, Application US/10004381
; Publication No. US2002015578A1
; GENERAL INFORMATION:
; APPLICANT: SZOSTAK, JACK W.
; APPLICANT: WILSON, DAVID S.
; APPLICANT: KEEFE, ANTHONY D.
; TITLE OF INVENTION: STREPTAVIDIN-BINDING PEPTIDES AND USES
; FILE REFERENCE: 00786/388002
; CURRENT APPLICATION NUMBER: US/10/004,381
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,541
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: selected peptide
US-10-004-381-11

Query Match 27.1%; Score 70; DB 13; Length 102;
Best Local Similarity 42.9%; Pred. No. 0.33;
Matches 18; Conservative 4; Mismatches 6; Indels 14; Gaps 3;

QY 1 MDEKTTGW--RGHVVVGLAGELEQLRARLEHHPOGOREPMM 40
DB 1 MDEK-TTHMRVYHHLADG-----LEQHPQGORRPLV 30

Search completed: July 15, 2004, 20:45:03
Job time : 45.2353 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:32:20 ; Search time 16.0353 Seconds
(without alignments)
151.318 Million cell updates/sec

Title: US-10-004-381-25

Perfect score: 258
Sequence: 1 MDEKTTGWMGHVVEGLAGE.....LEHHPGQREPMWGGCKXG 47

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backflieal.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	24.4	543	3 US-09-413-814-91	Sequence 91, Appl
2	63	24.4	544	3 US-09-413-814-81	Sequence 81, Appl
3	62	24.0	356	4 US-09-252-991A-19549	Sequence 19549, A
4	61	23.6	239	4 US-09-134-000C-5025	Sequence 5025, Ap
5	60	23.3	152	4 US-09-252-991A-18469	Sequence 18469, A
6	60	23.3	432	4 US-09-252-991A-21794	Sequence 21794, A
7	59.5	23.1	543	4 US-09-252-991A-29830	Sequence 29830, A
8	58.5	22.7	197	4 US-09-252-991A-24825	Sequence 24825, A
9	58	22.5	113	2 US-08-248-839C-6	Sequence 6, Appl1
10	58	22.5	207	4 US-09-328-352-5708	Sequence 5708, Ap
11	58	22.5	286	4 US-09-252-991A-27885	Sequence 27885, A
12	58	22.5	628	4 US-09-252-991A-22131	Sequence 22131, A
13	57.5	22.3	1065	3 US-09-412-545-2	Sequence 2, Appl1
14	57	22.1	296	4 US-09-252-991A-30737	Sequence 30737, A
15	57	22.1	495	4 US-08-311-731A-3	Sequence 3, Appl1
16	57	22.1	674	4 US-09-252-991A-18107	Sequence 18107, A
17	57	22.1	904	4 US-09-252-991A-29119	Sequence 29119, A
18	57	22.1	1136	4 US-09-252-991A-31394	Sequence 31394, A
19	56.5	21.9	220	4 US-09-252-991A-24796	Sequence 24796, A
20	56.5	21.9	283	4 US-09-252-991A-20071	Sequence 20071, A
21	56	21.7	468	4 US-09-199-637A-29	Sequence 29, Appl1
22	55.5	21.5	1008	4 US-09-252-991A-19329	Sequence 19329, A
23	55.5	21.5	1388	4 US-09-572-191-2	Sequence 2, Appl1
24	55.5	21.5	1388	4 US-09-723-262-2	Sequence 2, Appl1
25	55.5	21.5	1388	4 US-09-723-219-2	Sequence 2, Appl1
26	55	21.3	162	4 US-09-252-991A-20901	Sequence 20901, A
27	55	21.3	184	4 US-09-252-991A-30468	Sequence 30468, A

28	55	21.3	578	4 US-09-252-991A-31318	Sequence 31318, A
29	54.5	21.1	459	4 US-09-134-000C-6639	Sequence 6639, Ap
30	54.5	21.1	478	4 US-09-252-991A-31542	Sequence 31542, A
31	54.5	21.1	602	4 US-09-252-991A-30529	Sequence 30529, A
32	54.5	21.1	1841	2 US-08-804-227C-6	Sequence 6, Appl1
33	54.5	21.1	4630	4 US-09-091-609-2	Sequence 2, Appl1
34	54.5	21.1	5215	3 US-09-105-537-2	Sequence 2, Appl1
35	54	20.9	185	4 US-09-252-991A-19647	Sequence 19647, A
36	54	20.9	499	4 US-09-252-991A-27221	Sequence 27221, A
37	54	20.9	631	4 US-09-252-991A-26444	Sequence 26444, A
38	53.5	20.7	266	4 US-09-252-991A-24870	Sequence 24870, A
39	53.5	20.7	511	4 US-09-252-991A-18418	Sequence 18418, A
40	53.5	20.7	692	4 US-09-252-991A-19069	Sequence 19069, A
41	53	20.5	224	4 US-09-199-637A-175	Sequence 175, App
42	53	20.5	258	4 US-09-252-991A-17377	Sequence 17377, A
43	53	20.5	648	4 US-09-252-991A-23682	Sequence 23682, A
44	53	20.5	752	4 US-09-252-991A-17355	Sequence 17355, A
45	52.5	20.3	69	4 US-09-328-352-4132	Sequence 4132, Ap

ALIGNMENTS

```

RESULT 1
US-09-413-814-91
; Sequence 91, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bioecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hoelle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; EARLIER FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 91
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-91

Query Match      24.4% Score 63; DB 3; Length 543;
Best Local Similarity 35.1%; Pred. No. 2.1;
Matches 13; Conservative 6; Mismatches 10; Indels 8; Gaps 1;

QY      8 WRGSHVVEG-----LAGELEQRLARLEHHPGQGR 36
Db      235 WLSGEVLBEGLARVYTKLSGALRRARVPVDEHPARR 271

RESULT 2
US-09-413-814-81
; Sequence 81, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bioecker, Helmut
; APPLICANT: Brandt, Petra

```

```
/ APPLICANT: Cino, Paul M
/ APPLICANT: Dougherty, Brian A
/ APPLICANT: Goldberg, Steven L
/ APPLICANT: Hofle, Gerhard
/ APPLICANT: Mueller, Joachim
/ APPLICANT: Reichenbach, Hans
/ TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
/ FILE REFERENCE: PCT/US 99/23535
/ CURRENT APPLICATION NUMBER: US/09/413,814
/ CURRENT FILING DATE: 1999-10-07
/ EARLIER APPLICATION NUMBER: DE 198 46 493.2
/ PRIOR FILING DATE: 1998-10-09
/ NUMBER OF SEQ ID NOS: 107
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 81
/ LENGTH: 544
/ TYPE: PRT
/ ORGANISM: Sorangium cellulosum
/ US-09-413-814-81

Query Match
Best Local Similarity 35.1%; Score 63; DB 3; Length 544;
Matches 13; Conservative 6; Mismatches 10; Indels 8; Gaps 1;

OY 8 WRGSHVEG-----LAGELEQLRLEHHHPOGOR 36
Db 236 WLSGEVLGEELRYWTKLSGALRRARVPDPDHEPAGR 272

RESULT 3
US-09-252-991A-19549
/ Sequence 19549, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 19549
/ LENGTH: 356
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
/ US-09-252-991A-19549

Query Match
Best Local Similarity 41.0%; Score 62; DB 4; Length 356;
Matches 16; Conservative 5; Mismatches 12; Indels 6; Gaps 2;

OY 7 GWRGSHVEGLAGELEQLRLEHHHPO--GOREPMMSG 43
Db 183 GHRGSHAGAGEPRQ----RATBQHPKAVGROOQIABG 217

RESULT 4
US-09-134-000C-5025
/ Sequence 5025, Application US/09134000C
/ Patent No. 6617156
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Stamm et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
/ FILE REFERENCE: 032796-032
/ CURRENT APPLICATION NUMBER: US/09/134,000C
/ CURRENT FILING DATE: 1998-08-13
/ PRIOR APPLICATION NUMBER: US 60/055,778
```

```
/ PRIOR FILING DATE: 1997-08-15
/ NUMBER OF SEQ ID NOS: 6812
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 5025
/ LENGTH: 239
/ TYPE: PRT
/ ORGANISM: Enterococcus faecalis
/ US-09-134-000C-5025

Query Match
Best Local Similarity 48.5%; Score 61; DB 4; Length 239;
Matches 16; Conservative 3; Mismatches 10; Indels 4; Gaps 2;

OY 8 WRGSHVEGLAGELEQLRLEHHHPOGOREPMM 40
Db 34 W-GGEVITGLMASWEQERAKEHE--GRETMM 62

RESULT 5
US-09-252-991A-18469
/ Sequence 18469, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 18469
/ LENGTH: 152
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
/ US-09-252-991A-18469

Query Match
Best Local Similarity 36.8%; Score 60; DB 4; Length 152;
Matches 14; Conservative 5; Mismatches 19; Indels 0; Gaps 0;

OY 10 GSHVEGLAGELEQLRLEHHHPOGOREPMMSGGCKG 47
Db 74 GSHVEVAGLVEGYQLRTVDHRAGRPLAAGQLG 111

RESULT 6
US-09-252-991A-21794
/ Sequence 21794, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 21794
/ LENGTH: 432
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
/ FEATURE:
/ NAME/KEY: UNSURE
/ LOCATION: (405)
/ OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
```

US-09-252-991A-21794

Query Match 23.3%; Score 60; DB 4; Length 432;

Best Local Similarity 52.9%; Pred. No. 4.1;

Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 27 RLEHHHGGREPMWGG 43

Db 143 RLHHHGGRRQPVLAG 159

RESULT 7

US-09-252-991A-29830

; Sequence 29830; Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 29830

; LENGTH: 543

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-29830

Query Match 23.1%; Score 59.5; DB 4; Length 543;

Best Local Similarity 42.9%; Pred. No. 6.3;

Matches 15; Conservative 3; Mismatches 14; Indels 3; Gaps 2;

QY 7 GWRGHVVEGLAG--ELEQLRLLEHHHPC-GOREP 38

Db 183 GGSAGHAFQGLADLVLDQLGVACQRRHHPDMKHP 217

RESULT 8

US-09-252-991A-24825

; Sequence 24825; Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 24825

; LENGTH: 197

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-24825

Query Match 22.7%; Score 58.5; DB 4; Length 197;

Best Local Similarity 44.8%; Pred. No. 2.7;

Matches 13; Conservative 2; Mismatches 13; Indels 1; Gaps 1;

QY 8 WRGHVVEGLAGLEQLRLLEHHHPCG 35

Db 112 WRGGRAPGDPGRHRRRLRLRRHHHPCGR 140

RESULT 9

US-08-248-839C-6

; Sequence 6; Application US/08248839C

; Patent No. 5843702

; GENERAL INFORMATION:

; APPLICANT: McConnell, David

; APPLICANT: Devine, Kevin

; TITLE OF INVENTION: A Gene Expression System

; NUMBER OF SEQUENCES: 185

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 58437020 No. 58437020th America, Inc.

; STREET: 405 Lexington Avenue

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/248,839C

; FILING DATE: 25-MAY-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Gregg, Valela A.

; REGISTRATION NUMBER: 35,127

; REFERENCE/DOCKET NUMBER: 3614.214-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 113 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: Protein

US-08-248-839C-6

Query Match 22.1%; Score 58; DB 2; Length 113;

Best Local Similarity 40.0%; Pred. No. 1.7;

Matches 12; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY 9 RGHVVEGLAGLEQLRLLEHHHPCGOREP 38

Db 10 RGRKQDEIAGHIGVSRARYSHYENGSRBP 39

RESULT 10

US-09-328-352-5708

; Sequence 5708; Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; PRIOR FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 5708

; LENGTH: 207

; TYPE: PRT

; ORGANISM: Acinetobacter baumannii

US-09-328-352-5708

Query Match 22.5%; Score 58; DB 4; Length 207;

Best Local Similarity 45.5%; Pred. No. 3.3;

Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 22 EQLRLLEHHHPCGOREPMWGG 43

Db 106 QVLSRRFSHPPLCKMKPEMATRG 127

RESULT 11
US-09-252-991A-27885
; Sequence 27885, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27885
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27885

Query Match 22.5%; Score 58; DB 4; Length 286;
Best Local Similarity 33.3%; Pred. No. 4.8;
Matches 15; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

QY 3 EKTGRRGHVVEGLAEQLRLRLEHHPOGOREPMMSGGCKL 47
Db 78 EHTHDPGRSRRVAVAGGFVRQQGRLVDGAGNADPLLAGKGVG 122

RESULT 12
US-09-252-991A-22131
; Sequence 22131, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22131
; LENGTH: 628
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22131

Query Match 22.5%; Score 58; DB 4; Length 628;
Best Local Similarity 44.4%; Pred. No. 12;
Matches 16; Conservative 1; Mismatches 11; Indels 8; Gaps 1;

QY 9 RGGHVVVEGLAEQLRLRLEHHPOGOREPMMSGGCKL 36
Db 126 RGGHVVVEGLAEQLRLRLEHHPOGOREPMMSGGCKL 161

RESULT 13
US-09-412-545-2
; Sequence 2, Application US/09412545
; Patent No. 6255095
; GENERAL INFORMATION:
; APPLICANT: Prescott, Stephen M.
; APPLICANT: Ding, Li

; APPLICANT: Traer, Elie
; TITLE OF INVENTION: HUMAN DIACYLGLYCEROL KINASE IOTA
; FILE REFERENCE: 1321.2.25
; CURRENT APPLICATION NUMBER: US/09/412,545
; CURRENT FILING DATE: 1999-10-05
; EARLIER APPLICATION NUMBER: 60/103,079
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1065
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-412-545-2

Query Match 22.3%; Score 57.5; DB 3; Length 1065;
Best Local Similarity 25.9%; Pred. No. 26;
Matches 14; Conservative 9; Mismatches 20; Indels 11; Gaps 1;

QY 4 KTTGRRGH-----VVEGLAEQLRLRLEHHPOGOREPMMSGGCKL 46
Db 469 KTNMGGITDEPRVSKILQVEBDGTVVOLDKRWNLHVERNPDLPELEDGVCKL 522

RESULT 14
US-09-252-991A-30737
; Sequence 30737, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30737
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30737

Query Match 22.1%; Score 57; DB 4; Length 296;
Best Local Similarity 46.2%; Pred. No. 6.9;
Matches 12; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 18 AGELEQLRLRLEHHPOGOREPMMSGG 43
Db 55 AGVLDLRLVAAEHHRMLRGQFQVGG 80

RESULT 15
US-08-311-731A-3
; Sequence 3, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LARAE FOR
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210


```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 495 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIUM TUBERCULOSIS
;
US-08-311-731A-3

```

```

QY      12  HVYEGAGELBOLRLRLEHHHPOGGRP--MMSGGCK 45
      |  ::  :| ||| ||| ||| ||| |||  ::  |||
DB      22  HYKKVAVBLDTRRLRLREYGRATEPVAVVGIGCR 57

```

Search completed: July 15, 2004, 20:37:00
 Job time : 17.0353 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:30:55 ; Search time 29.5059 Seconds

(without alignments)
406.349 Million cell updates/sec

Title: US-10-004-381-25_COPY_1_38

Perfect score: 208
Sequence: 1 MDEKTGMGRGHVVEGLAGELBGLRLRLHHPGQREPP 38

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	66.5	32.0	739	16	Q82BY9 streptomyc
2	65.5	31.5	189	2	Q8KRD7
3	61	29.3	773	2	033541
4	60	28.8	588	16	Q8DH85
5	59	28.4	139	4	Q96MD1
6	59	28.4	342	16	Q8XAY8
7	59	28.4	342	16	Q83RDB
8	59	28.4	1083	3	094189
9	59	28.4	2360	3	094188
10	58.5	28.1	274	16	Q8P4G4
11	58	27.9	333	4	Q81UY4
12	58	27.9	491	2	Q9ZCD9
13	58	27.9	1161	16	Q9J0L0
14	58	27.9	1181	16	Q82DQ5
15	57	27.4	229	16	Q7U554
16	57	27.4	489	16	Q8FLJ7

17	57	27.4	663	16	Q91SD9	Q915d9 pseudomonas
18	56.5	27.2	274	2	Q83V70	Q83v70 halomonas s
19	56.5	27.2	377	16	Q9RKG2	Q9Rkg2 streptomyc
20	56.5	27.2	1559	4	Q7Z5S5	Q7z5s5 homo sapien
21	56	26.9	600	2	Q7WUL9	Q7wul9 pseudomonas
22	56	26.9	717	11	Q8BWM4	Q8bwm4 mus musculu
23	56	26.9	976	5	Q9U158	Q9u158 leishmania
24	55.5	26.7	290	16	Q82ZC6	Q82zc6 enterococcu
25	55.5	26.7	360	16	Q82ZM7	Q82zm7 pseudomonas
26	55.5	26.7	430	5	Q9W4Q1	Q9w4q1 dirosophila
27	55.5	26.7	440	16	Q8K692	Q8k692 streptococc
28	55.5	26.7	442	16	Q879H3	Q879h3 streptococc
29	55.5	26.7	1388	4	Q9NS87	Q9ns87 homo sapien
30	55	26.4	134	4	Q8WVX5	Q8wvx5 homo sapien
31	55	26.4	453	16	Q88LX2	Q88lx2 pseudomonas
32	55	26.4	487	13	Q8JGR1	Q8jgr1 brachydanio
33	55	26.4	487	13	Q7ZU28	Q7zu28 brachydanio
34	55	26.4	605	4	Q96NPF	Q96nfp homo sapien
35	55	26.4	714	16	Q89UN0	Q89un0 bradyrhicob
36	55	26.4	1206	10	Q7XNE8	Q7xne8 oryza sativ
37	54.5	26.2	536	5	Q8T9S0	Q8t9s0 aplysia cal
38	54.5	26.2	536	5	Q8T0Y9	Q8t0y9 aplysia cal
39	54.5	26.2	786	16	Q7V5N6	Q7v5n6 prochloroco
40	54.5	26.2	835	10	Q81H08	Q81h08 oryza sativ
41	54	26.0	159	16	Q82J34	Q82j34 streptomyc
42	54	26.0	321	16	Q9A200	Q9a200 caulobacter
43	54	26.0	329	10	Q8L6S5	Q8l6s5 oryza sativ
44	54	26.0	469	16	Q8PRC0	Q8prc0 xanthomonas
45	54	26.0	674	11	Q60979	Q60979 mus musculu

ALIGNMENTS

RESULT 1	ID	Q82BY9	PRELIMINARY;	PRT;	739 AA.
AC	Q82BY9	01-JUN-2003 (TREMBLrel_24, Created)			
DT	01-JUN-2003 (TREMBLrel_24, Last sequence update)				
DT	01-OCT-2003 (TREMBLrel_25, Last annotation update)				
DE	Putative regulatory protein.				
GN	SAV5565.				
OS	Streptomyces avermitilis.				
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
OC	Streptomycineae; Streptomycetaceae; Streptomycetes.				
OX	NCBI_TaxId=33903;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;				
RX	MEDLINE=21477403; PubMed=11572948;				
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,				
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,				
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;				
RT	"Genome sequence of an industrial microorganism Streptomyces				
RT	avermitilis: deducing the ability of producing secondary				
RT	metabolites.";				
RT	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).				
[2]					
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;				
RX	MEDLINE=22608306; PubMed=12692562;				
RA	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,				
RA	Sakaki Y., Hattori M., Omura S.;				
RT	"Complete genome sequence and comparative analysis of the industrial				
RT	microorganism Streptomyces avermitilis.";				
RT	Nat. Biotechnol. 21:526-531 (2003).				
DR	EMBL; AP005043; BAC7277.1; .				
DR	GO; GO:0005524; F:ATP binding; IEA.				
DR	GO; GO:0003824; F:catalytic activity; IEA.				
DR	InterPro; IPR003594; A:ATPbind ATPase.				
DR	InterPro; IPR001932; P:PC-1like.				
DR	Pfam; PF02518; HATPase_c; 1.				

```

DR SMART: SM00331; PP2C_SIG; 1.
KM Complete proteome.
SQ SEQUENCE 739 AA; 77478 MW; 63853F93EC7A2B6 CRC64;

Query Match
Best Local Similarity 32.0%; Score 66.5; DB 16; Length 739;
Matches 14; Conservative 5; Mismatches 13; Indels 11; Gaps 1;

QY 1 MDEKTTGMRG-----GHVVGGLAGLEQLPARLEHHP 32
DB 341 LDEATGWRGPETGMAPGPRLARVWHSSENRLEDRALRHP 383

RESULT 2
Q8KRD7 PRELIMINARY; PRT; 189 AA.
AC Q8KRD7;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Transcription repressor KfrA-like protein.
OS Rhodococcus erythropolis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MP50;
RA MEDLINE=22083458; PubMed=12089004;
RA Troit S., Burger S., Calaminius C., Stolz A.;
RT "Cloning and Heterologous Expression of an Eranthisselective Amidase
from Rhodococcus erythropolis Strain MP50."
RL Appl. Environ. Microbiol. 68:3279-3286(2002).
DR EMBL; AY026386; AKL1725.1; -.
SQ SEQUENCE 189 AA; 20101 MW; 4B8B5D00A093FFFD CRC64;

Query Match
Best Local Similarity 31.5%; Score 65.5; DB 2; Length 189;
Matches 20; Conservative 2; Mismatches 13; Indels 9; Gaps 1;

QY 3 EKTGWRGSHV-----VEGLAGLEQLPARLEHHPGORE 37
DB 104 EQHTAERDGAVTARDRALGEVEGLRYELRQLRGLETARADARE 147

RESULT 3
Q33541 PRELIMINARY; PRT; 773 AA.
ID Q33541;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Sensor protein RsaA.
GN RsaA.
OS endosymbiont of Rickettsia pachyptila.
OC Bacteria; Proteobacteria; Gammaproteobacteria;
OC sulfur-oxidizing symbionts.
OX NCBI_TaxID=54396;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97438512; PubMed=9293000;
RA Hughes D.S., Felbeck H., Stein J.L.;
RT "A histidine protein kinase homolog from the endosymbiont of the
hydrothermal vent tubeworm Rickettsia pachyptila."
RL Appl. Environ. Microbiol. 63:3494-3498(1997).
CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
KINASES.
DR EMBL; U93704; AAB71131.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.

```

```

DR GO; GO:0000156; F:two-component response regulator activity; IEA.
DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR GO; GO:0007600; F:sensory perception; IEA.
DR GO; GO:0000160; F:two-component signal transduction system (p. . .); IEA.
DR InterPro; IPR003594; Actbind ATPase.
DR InterPro; IPR004358; Bact_sens_pr_C.
DR InterPro; IPR005467; His_Kinase.
DR InterPro; IPR003661; His_kina_N.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF02518; HATPase_1.
DR Pfam; PF00512; HisKA_1.
DR PRINTS; PR00072; response_reg; 1.
DR PRODOM; PD000039; BCTRSENSOR.
DR SMART; SM00387; HATPase_c; 1.
DR SMART; SM00448; REC; 1.
DR SMART; SM00388; HisKA_1.
DR PROSITE; PS50109; HIS_KIN; 1.
DR PROSITE; PS50110; RESPONSE_REGULATORY; 1.
KW Kinase; Phosphorylation; Sensory transduction; Transferase.
SQ SEQUENCE 773 AA; 87086 MW; 2C5643B64F373B27 CRC64;

Query Match
Best Local Similarity 29.3%; Score 61; DB 2; Length 773;
Matches 13; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 8 WRGSHVVEGLAGLEQLPARLEHHPGORE 37
DB 346 WRGPREIELEKLTALAEYHAHHTQORE 375

RESULT 4
Q8DH85 PRELIMINARY; PRT; 588 AA.
ID Q8DH85;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Penicillin-binding protein.
GN TLR2074.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsunoto M., Matsuno A., Nakazaki N.,
RA Shingo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.,
RT "Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1."
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005376; BAC09626.1; -.
DR GO; GO:0008658; F:penicillin binding; IEA.
DR GO; GO:0009227; P:cell wall biosynthesis (sensu Bacteria); IEA.
DR InterPro; IPR005311; PBP_dimer.
DR InterPro; IPR001460; Transpeptidase.
DR Pfam; PF03717; PBP_dimer; 1.
DR Pfam; PF00905; Transpeptidase; 1.
KW Complete proteome.
SQ SEQUENCE 588 AA; 65048 MW; 9E24A8544A102719 CRC64;

Query Match
Best Local Similarity 28.8%; Score 60; DB 16; Length 588;
Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 12 HVEGLAGLEQLPARLEHHP 32
DB 442 HVTIDGLYDENGLOLQRLRHP 462

```

```

RESULT 5
O96MD1 PRELIMINARY; PRT; 139 AA.
AC O96MD1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ32575.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=SpLeen;
RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Otsuki T., Saito H., Oca T., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamaashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Magatsuuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isegai T.;
RT "MEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK057137; BAB71369.1; -.
KW Hypothetical protein.
SQ SEQUENCE 139 AA; 15014 MW; 3A59D86197848F19 CRC64;

Query Match 28.4%; Score 59; DB 4; Length 139;
Best Local Similarity 44.8%; Pred. No. 11;
Matches 13; Conservative 3; Mismatches 11; Indels 2; Gaps 1;

Oy 3 EKTGWRGHVGEIAGELBQLRLRRL 31
Db 92 ERATGMRLSHEVSGSGQLQATSCR-HH 118

RESULT 6
O8XAY8 PRELIMINARY; PRT; 342 AA.
AC O8XAY8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative transpore system permease protein.
GN YDEY OR Z2191 OR EGS2121.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., Ili, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Groetbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kohara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
DR EMBL; AE005353; AAG56252.1; -.

```

```

DR EMBL; AP002557; BAB35544.1; -.
DR PIR; A90894; A90894.
DR PIR; H85723; H85723.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001851; Bac_inmem_transp.
DR Pfam; PF02653; BPD_transp_2; 1.
KW Complete proteome.
SQ SEQUENCE 342 AA; 36365 MW; 6953A23C07C305A8 CRC64;

Query Match 28.4%; Score 59; DB 16; Length 342;
Best Local Similarity 57.1%; Pred. No. 29;
Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Oy 8 WRGKRVGEIAGELBQLRLRRL 28
Db 135 WTGKRVGEIAGELBQLRLRRL 155

RESULT 7
O83RD8 PRELIMINARY; PRT; 342 AA.
AC O83RD8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative transpore system permease protein.
GN YDEY OR SF1584 OR S1710.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang P., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441 (2002).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G., Ili, Rose D.J., Darling A.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786 (2003).
DR EMBL; AE015180; FAN43171.1; -.
DR EMBL; AE015983; AAP17063.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001851; Bac_inmem_transp.
DR Pfam; PF02653; BPD_transp_2; 1.
KW Complete proteome.
SQ SEQUENCE 342 AA; 36396 MW; E5FE60B2AA54604B CRC64;

Query Match 28.4%; Score 59; DB 16; Length 342;
Best Local Similarity 57.1%; Pred. No. 29;
Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Oy 8 WRGKRVGEIAGELBQLRLRRL 28
Db 135 WTGKRVGEIAGELBQLRLRRL 155

```

```
RESULT 8
ID 094189 PRELIMINARY; PRT; 1083 AA.
AC 094189;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Phosphatidylinositol 3-kinase TOR1 (Fragment).
GN TOR1.
OS Cryptococcus neoformans var. neoformans.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC Tremellulomycetidae; Tremellales; Tremellaceae; Filobasidiella.
OX NCBI_TaxId=40410;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B3501;
RX MEDLINE=9262981; PubMed=10330150;
RA Cruz M.C., Cavallo L.M., Goriach J.M., Cox G., Perfect J.R.,
RA Cardenas M.B., Heltman J.;
RT Rapamycin antifungal action is mediated via conserved complexes with
RT FKBP12 and TOR kinase homologs in Cryptococcus neoformans.";
RL Mol. Cell. Biol. 19:4101-4112(1999).
DR EMBL; AF098973; AAD16274.1; -.
DR HSSP; PA2345; 1AUE.
DR GO; GO:0004428; F:inositol/phosphatidylinositol kinase activity; IEA.
DR InterPro; IPR003151; FAT.
DR InterPro; IPR003152; FATC.
DR InterPro; IPR004003; P13_P14_kinase.
DR InterPro; IPR008941; TPR1-like.
DR Pfam; PF02259; FAT; 1.
DR Pfam; PF02260; FATC; 1.
DR Pfam; PF00454; P13_P14_kinase; 1.
DR SMART; SM00146; PI3Kc; 1.
DR PROSITE; PS00915; P13_4_KINASE_1; 1.
DR PROSITE; PS50290; P13_4_KINASE_3; 1.
KW Kinase.
FT NON_TER
FT NON_TER
SQ SEQUENCE 1083 AA; 123973 MW; 0C48A17D758353A9 CRC64;

Query Match 28.4%; Score 59; DB 3; Length 1083;
Best Local Similarity 35.1%; Pred. No. 1e+02;
Matches 13; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVEGLAGLEQLRRLRLEHHHPCGORE 37
DB 581 LEEASKHYFGDHDIPGMLGVLEPLHRIEIVENGPOTLRE 617

RESULT 9
ID 094188 PRELIMINARY; PRT; 2360 AA.
AC 094188;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Phosphatidylinositol 3-kinase TOR1.
GN TOR1.
OS Cryptococcus neoformans var. neoformans.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC Tremellulomycetidae; Tremellales; Tremellaceae; Filobasidiella.
OX NCBI_TaxId=40410;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H99;
RX MEDLINE=9262981; PubMed=10330150;
RA Cruz M.C., Cavallo L.M., Goriach J.M., Cox G., Perfect J.R.,
RA Cardenas M.B., Heltman J.;
RT Rapamycin antifungal action is mediated via conserved complexes with
RT FKBP12 and TOR kinase homologs in Cryptococcus neoformans.";
RL Mol. Cell. Biol. 19:4101-4112(1999).
```

```
DR EMBL; AF098972; AAD16273.1; -.
DR HSSP; PA2345; 1AUE.
DR GO; GO:0004428; F:inositol/phosphatidylinositol kinase activity; IEA.
DR InterPro; IPR003151; FAT.
DR InterPro; IPR003152; FATC.
DR InterPro; IPR004003; P13_P14_kinase.
DR Pfam; PF02259; FAT; 1.
DR Pfam; PF02260; FATC; 1.
DR Pfam; PF00454; P13_P14_kinase; 1.
DR SMART; SM00146; PI3Kc; 1.
DR PROSITE; PS00915; P13_4_KINASE_1; 1.
DR PROSITE; PS50290; P13_4_KINASE_3; 1.
KW Kinase.
SQ SEQUENCE 2360 AA; 267216 MW; ED4A1059B1AA2B2A CRC64;

Query Match 28.4%; Score 59; DB 3; Length 2360;
Best Local Similarity 35.1%; Pred. No. 2.4e+02;
Matches 13; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 1 MDEKTTGMRGHHVEGLAGLEQLRRLRLEHHHPCGORE 37
DB 1858 LEEASKHYFGDHDIPGMLGVLEPLHRIEIVENGPOTLRE 1894

RESULT 10
ID 08P4G4 PRELIMINARY; PRT; 274 AA.
AC 08P4G4;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Lipopolysaccharide core biosynthesis glycosyl transferase
DE (Glycosyltransferase).
GN XCG3744 OR WAXE.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxId=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaglio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camavau J., Cardoso J., Chambergo F., Clapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis U., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Teai S.M., White F.F.,
RA Setubal J.C., Kiteajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=8004;
RX "Identification and Cloning of a Novel Gene Involved in EPS
RT Biosynthesis of Xanthomonas campestris pv. campestris.";
RL Submitted (JUN-2003) to the EMBL/Genbank/DBJ databases.
DR EMBL; AE012495; AAM33001.1; -.
DR EMBL; AY329625; AAP33588.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR001173; Glyco_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Transferase; Complete proteome.
```

SQ SEQUENCE 274 AA; 31016 MW; 28F541FA102D64B CRC64;
 Query Match 28.1%; Score 58.5; DB 16; Length 274;
 Best Local Similarity 38.2%; Pred. No. 26;
 Matches 13; Conservative 3; Mismatches 15; Indels 3; Gaps 1;
 QY 2 DEKTTGMRGSG--HVEGLAGELEQURARLEHHP 32
 DB 144 DRRGGGGRGKEIHEASVDGTATLRGDLIHP 177
 RESULT 11
 Q81U04 PRELIMINARY; PRT; 333 AA.
 ID 081U04;
 AC 081U04;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypochemical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RA Strausberg R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC040438; AAH0438.1; -.
 DR InterPro; IPR001073; C1q.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR008983; TNF_like.
 DR Pfam; PF00386; C1q; 1.
 DR Pfam; PF01381; Collagen; 3.
 DR PRINTS; PR00007; COMPLEMENTC1Q.
 DR SMART; SM00110; C1Q; 1.
 DR PROSITE; PS01113; C1Q; 1.
 DR KMW
 KW Hypochemical protein.
 SQ SEQUENCE 333 AA; 34649 MW; E9EBF803A034E307 CRC64;
 Query Match 27.9%; Score 58; DB 4; Length 333;
 Best Local Similarity 40.0%; Pred. No. 37;
 Matches 14; Conservative 6; Mismatches 15; Indels 0; Gaps 0;
 QY 2 DEKTTGMRGSHVVEGLAGELEQURARLEHHPGOR 36
 DB 90 DQSGRSGPKRGKGLAGPWEKGLRGETGPOGOK 124
 RESULT 12
 Q9ZGD9 PRELIMINARY; PRT; 491 AA.
 ID 09ZGD9;
 AC 09ZGD9;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Oxygenase homolog.
 OS Streptomyces cyanogenus.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomyces; Streptomycetaceae; Streptomyces.
 NC NCB1_TaxID=80860;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S136;
 RC MEDLINE=99132695; PubMed=9933932;
 RA Westrich L., Domann S., Faust B., Bedford D., Hopwood D.A.,
 RA Bechthold A.;
 RT "Cloning and characterization of a gene cluster from Streptomyces
 cyanogenus S136 probably involved in landomycin biosynthesis.";
 RL FEMS Microbiol. Lett. 170:381-387 (1999).
 DR EMBL; AF080235; AAD13534.1; -.
 DR GO; GO:0004497; F:monooxygenase activity; IEA.

DR GO; GO:0006725; P:aromatic compound metabolism; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR000733; P:monooxygenase.
 DR InterPro; IPR002938; Mox_FAD_binding.
 DR InterPro; IPR003025; NAD_BS.
 DR InterPro; IPR003042; Rng_moxxygenase.
 DR Pfam; PF01494; FAD binding 3; 1.
 DR Pfam; PF01360; Monooxygenase; 1.
 DR PRINTS; PR00420; RINGMOXGNASE.
 SQ SEQUENCE 491 AA; 52062 MW; 096FD4A642275941 CRC64;
 Query Match 27.9%; Score 58; DB 2; Length 491;
 Best Local Similarity 40.5%; Pred. No. 57;
 Matches 15; Conservative 4; Mismatches 14; Indels 4; Gaps 2;
 QY 2 DEKTTGMRGSHVVEGLAGE--LEQURARLEHHPGOR 36
 DB 293 DSVNLGWLKLAAYVNGIAGLIDSYHS--ERHPVGR 327
 RESULT 13
 Q910L0 PRELIMINARY; PRT; 1161 AA.
 ID 0910L0;
 AC 0910L0;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE DNA-directed RNA polymerase beta chain.
 GN RPOB OR SCO4654 OR SCD82.26.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomyces; Streptomycetaceae; Streptomyces.
 NC NCB1_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RC MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96 (1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RC MEDLINE=2196410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser T., Lake L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147 (2002).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 SUBSTRATES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE +
 {RNA} (N).

```

CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
DR EMBL: A939121, CAB7428.1; -.
DR HSSP: Q9KWD7, 1HOM.
DR GO: GO:0003677, F:DNA binding; IEA.
DR GO: GO:0003899, F:DNA-directed RNA polymerase activity; IEA.
DR GO: GO:0016740, F:transcriptase activity; IEA.
DR GO: GO:0006350, P:transcription; IEA.
DR InterPro: IPR007121, RNA_pol_B.
DR InterPro: IPR007644, RNA_pol_Rpb2_1.
DR InterPro: IPR007642, RNA_pol_Rpb2_2.
DR InterPro: IPR007645, RNA_pol_Rpb2_3.
DR InterPro: IPR007120, RNA_pol_Rpb2_6.
DR InterPro: IPR007641, RNA_pol_Rpb2_7.
DR Pfam: PF04563, RNA_pol_Rpb2_1; 1.
DR Pfam: PF04561, RNA_pol_Rpb2_2; 2.
DR Pfam: PF04565, RNA_pol_Rpb2_3; 1.
DR Pfam: PF00562, RNA_pol_Rpb2_6; 1.
DR Pfam: PF04560, RNA_pol_Rpb2_7; 1.
DR PROSITE: PS01166, RNA_POL_BETA, 1.
DR DNA-directed RNA polymerase; Transcription; Transferase;
KW Complete proteome.
SQ SEQUENCE 1161 AA; 128494 MW; A9C27B928E8191A8 CRC64;

Query Match 27.9%; Score 58; DB 16; Length 1161;
Best Local Similarity 45.5%; Pred. No. 1.5e+02;
Matches 15; Conservative 3; Mismatches 13; Indels 2; Gaps 2;

Qy 4 KTTGMRGHHVEGLAGELEQLRRLRLE-HHPQCG 35
Db 227 KALGWTTEQLLEEF-GEYESMRATLEKDHQCG 258

RESULT 14
Q82DQ5 PRELIMINARY; PRT; 1181 AA.
ID 082DQ5;
AC Q82DQ5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative RNA polymerase beta subunit.
GN RPOB OR SAV4914.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
NCBI_TaxId=33903;
RX STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites."
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis."
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL: APO05040; BAC72626.1; -.
DR GO: GO:0003677, F:DNA binding; IEA.
DR GO: GO:0003899, F:DNA-directed RNA polymerase activity; IEA.
DR GO: GO:0006350, P:transcription; IEA.
DR InterPro: IPR007121, RNA_pol_B.

```

```

DR InterPro: IPR007644, RNA_pol_Rpb2_1.
DR InterPro: IPR007642, RNA_pol_Rpb2_2.
DR InterPro: IPR007645, RNA_pol_Rpb2_3.
DR InterPro: IPR007120, RNA_pol_Rpb2_6.
DR InterPro: IPR007641, RNA_pol_Rpb2_7.
DR Pfam: PF04563, RNA_pol_Rpb2_1; 1.
DR Pfam: PF04561, RNA_pol_Rpb2_2; 1.
DR Pfam: PF04565, RNA_pol_Rpb2_3; 1.
DR Pfam: PF00562, RNA_pol_Rpb2_6; 1.
DR Pfam: PF04560, RNA_pol_Rpb2_7; 1.
DR PROSITE: PS01166, RNA_POL_BETA, 1.
KW Complete proteome.
SQ SEQUENCE 1181 AA; 130439 MW; AEBF77AC879ED8B CRC64;

Query Match 27.9%; Score 58; DB 16; Length 1181;
Best Local Similarity 45.5%; Pred. No. 1.5e+02;
Matches 15; Conservative 3; Mismatches 13; Indels 2; Gaps 2;

Qy 4 KTTGMRGHHVEGLAGELEQLRRLRLE-HHPQCG 35
Db 247 KALGWTTEQLLEEF-GEYESMRATLEKDHQCG 278

RESULT 15
Q7U554 PRELIMINARY; PRT; 229 AA.
ID 07U554;
AC 07U554;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SYN1855.
OS Synecchococcus sp. (strain WH8102).
OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.
NCBI_TaxId=84588;
RX MEDLINE=22825697; PubMed=12917641;
RA Palenik B., Brakhamsa B., Larimer F.W., Land M., Hauser L.,
RA Lamerdin J., Regala W., Allen E.E., McCarren J., Paulsen I.,
RA Dufresne A., Patensky F., Webb E.A., Waterbury J.;
RT "The genome of a motile marine Synecchococcus."
RL Nature 424:1037-1042(2003).
DR EMBL: BX569694; CAE08370.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 229 AA; 24435 MW; 851E411B69B319CB CRC64;

Query Match 27.4%; Score 57; DB 16; Length 229;
Best Local Similarity 37.5%; Pred. No. 33;
Matches 12; Conservative 4; Mismatches 16; Indels 0; Gaps 0;

Qy 5 TTGMRGHHVEGLAGELEQLRRLRLE-HHPQCG 36
Db 81 TSGMREIQAWVSAGADYALDATORHREGEOR 112

Search completed: July 15, 2004, 20:35:48
Job time : 30.5059 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:30:05 ; Search time 7.15294 Seconds
(without alignments)
276.623 Million cell updates/sec

Title: US-10-004-381-25_COPY_1_38
Perfect score: 208
Sequence: 1 MDEKTGMWGSHVVEGLAGELBQLRLARLEHHHRCGRFP 38

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60.5	29.1	177	1 HSLV_RHIL0	Q98C88 rhizobium 1
2	59	28.4	342	1 YDEY_ECOLI	P77672 escherichia
3	57	27.4	474	1 TRPC_CORGL	P06560 corynebacte
4	56.5	27.2	404	1 KVB3_HUMAN	O43448 homo sapien
5	56.5	27.2	1560	1 SMCK_HUMAN	P41229 homo sapien
6	55.5	26.7	188	1 HSLV_CAUCR	Q9A239 caulobacter
7	54	26.0	376	1 ALG_STRCO	Q9K434 streptomyce
8	53.5	25.7	404	1 KVB3_RAT	O63494 rattus norv
9	53	25.5	348	1 PLSX_SYNY3	P73950 synecocyst
10	53	25.5	589	1 IPEB_ASCSU	P23731 ascaris suu
11	52.5	25.2	184	1 HSLV_BRUME	O8Y311 bruceella me
12	52.5	25.2	184	1 HSLV_BRUSU	O8Y311 bruceella su
13	52	25.0	113	1 XRE_BACSU	P23769 bacillus su
14	52	25.0	356	1 GMB2_CAEEL	Q20636 caenorhabdi
15	52	25.0	499	1 AINX_HUMAN	O16352 homo sapien
16	52	25.0	504	1 AINX_MOUSE	P46660 mus musculu
17	52	25.0	505	1 AINX_RAT	P23565 rattus norv
18	52	25.0	2337	1 TOR2_SCHPO	Q9Y762 schizosacch
19	51.5	24.8	181	1 COB2_ECOLI	P46866 escherichia
20	51.5	24.8	273	1 GS80_BACSU	P08078 bacillus su
21	51.5	24.8	401	1 KVB1_MOUSE	O63277 mus musculu
22	51.5	24.8	408	1 KVB1_MOUSE	O28558 mustela put
23	51.5	24.8	419	1 KVB1_HUMAN	O14772 homo sapien
24	51.5	24.8	419	1 KVB1_RABIT	Q9XCT1 oryctolagus
25	51.5	24.8	465	1 FXD3_MOUSE	O61060 mus musculu
26	51	24.5	83	1 RS28_SULSO	Q98065 sulfolobus
27	51	24.5	261	1 YEVE_YEAST	P40078 saccharomyc
28	51	24.5	358	1 AROB_AERPE	Q9VE19 aeropyrum p
29	51	24.5	429	1 NPX2_MOUSE	O70340 mus musculu
30	51	24.5	431	1 NPX2_HUMAN	P47792 homo sapien
31	51	24.3	149	1 RL13_THEMA	Q9X135 thermotoga
32	50.5	24.3	176	1 HSLV_RHIME	Q92169 rhizobium m
33	50.5	24.3	176	1 HSLV_RHIME	Q92169 rhizobium m

34	50.5	24.3	1003	1 MBD6_HUMAN	Q9646 homo sapien
35	50	24.0	259	1 YAE2_PYPAB	Q9UC14 pyrococcus
36	50	24.0	401	1 KVB3_XENLA	O90M4 xenopus lae
37	50	24.0	473	1 RBL1_RHOCA	O32740 rhodobacter
38	49.5	23.8	366	1 ADH_ALCEU	P14940 alcaligenes
39	49.5	23.8	410	1 AAZA_MOUSE	O60613 mus musculu
40	49.5	23.8	176	1 DEPI_BORPE	O7WQ0 bordetella
41	49	23.6	176	1 DER2_BORPA	O7WQ0 bordetella
42	49	23.6	241	1 YJ03_ECOLI	P39403 escherichia
43	49	23.6	407	1 PEPT_STRP3	O8K802 streptococc
44	49	23.6	407	1 PEPT_STRP8	O8PH9 streptococc
45	49	23.6	407	1 PEPT_STRPY	Q9A0F4 streptococc

ALIGNMENTS

RESULT 1	ID	HSIV_RHIL0	STANDARD	PRT	177 AA.
AC	Q98C88				
DT	28-FEB-2003 (Rel. 41, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	ATP-dependent protease hslv (EC 3.4.25.-)				
GN	HSIV OR ML5007				
OS	Rhizobium loti (Mesorhizobium loti)				
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;				
OX	Phyllobacteriaceae; Mesorhizobium.				
NCBI_TaxID=381;					
NCBI [1]					
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MAF303099;				
RX	MEDLINE=21082950; PubMed=11214968;				
RA	Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,				
RA	Matanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,				
RA	Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,				
RA	Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,				
RA	Takeuchi C., Yamada M., Tabata S.,				
RT	"Complete genome structure of the nitrogen-fixing symbiotic bacterium				
RT	Mesorhizobium loti."				
RL	DNA Res. 7:331-338 (2000).				
CC	-1- FUNCTION: Protease subunit of a proteasome-like degradation				
CC	complex (By similarity).				
CC	-1- SUBUNIT: A double ring-shaped homohexamer of hslv is capped on				
CC	each side by a ring-shaped hslu homohexamer (By similarity).				
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).				
CC	-1- SIMILARITY: Belongs to peptidase family T1B. Hslv subfamily.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
CC	EMBL; AP003005; BAB5153.1; -				
DR	MEROPS; T01.006; -				
DR	HAMAP; MF_00248; -				
DR	InterPro; IPR001353; Peptidase_T1.				
DR	Pfam; PF00227; proteasome; 1.				
KW	Hydrolase; Protease; Threonine protease; Complete proteome.				
FT	ACT SITE				
FT	SEQUENCE				
FT	177 AA; 18986 MW; 62F3D3DC6ACBSF05 CRC64;				
QY	Query Match				
QY	Best Local Similarity				
QY	Matches				
QY	13; Conservative				
QY	6; Mismatches				
QY	5; Indels				
QY	7; Gaps				
QY	1;				
QY	9 RGHVVEGLAGE-----LEQLRLARLEHHP 32				
QY	42 KGGNVLAGPAGATADAFLLERLEATLEQYP 72				

RESULT 2
YOEY ECOLI STANDARD; PRT; 342 AA.
ID YOEY ECOLI P7672;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical ABC transporter permease protein ydey.
YOEY OR B1514.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OK NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12; MEDLINE=97251357; PubMed=9097039;
RX Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampei K., Seki Y., Sivasubraman S.,
RA Yamamoto Y., Horinouchi T.;
RA "A 570-bp DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Ref. 3363-337(1986)
CC -1- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM EGO/YDEYZ. PROBABLY RESPONSIBLE FOR THE TRANSLOCATION OF
CC THE SUBSTRATE ACROSS THE MEMBRANE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- SIMILARITY: Belongs to the binding-protein-dependent transport
CC system permease family. Arah/tpsc subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000249; AAC74587.1; -;
DR EMBL: D90794; BAA15201.1; -;
DR PIR: E64905; E64905.
DR Ecogene; Egl3807; ydey.
DR InterPro; IPR001851; Bac_inmem_transp.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF02653; BPD_transp_2; 1.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 14 34
FT TRANSMEM 19 59
FT TRANSMEM 70 90
FT TRANSMEM 93 113
FT TRANSMEM 115 135
FT TRANSMEM 157 177
FT TRANSMEM 213 233
FT TRANSMEM 252 272
FT TRANSMEM 284 304
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.

SQL SEQUENCE 342 AA; 36394 MW; 419E505026ABDE33 CRC64;
Query Match 28.4%; Score 59; DB 1; Length 342;
Best Local Similarity 57.1%; Pred. NO. 2.9;
Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 8 WRGHVVEGLAGELEQRLRL 28
DB 135 WTGGKMGLEPAELKQSLAPL 155
RESULT 3
TPSC CORGL STANDARD; PRT; 474 AA.
ID TPSC CORGL P06560;
DT 01-JAN-1998 (Rel. 06, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tryptophan biosynthesis protein trpC (includes: indole-3-glycerol
DE phosphate synthase (EC 4.1.1.48) (IGPS); N-(5'-phospho-
DE ribosyl)anthranilate isomerase (EC 5.3.1.24) (PRAI)].
GN TPSC OR CGJ3033.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteriales; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OK NCBI_Taxid=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=87117512; PubMed=3808947;
RX Matsui K., Sano K., Ohtsubo E.;
RT "Complete nucleotide and deduced amino acid sequences of the
RL Nucleic Acids Res. 14:10113-10114(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RX Nakagawa S.;
RT Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
RL -1- FUNCTION: Bifunctional enzyme that catalyzes two sequential steps
CC of tryptophan biosynthetic pathway. The first reaction is
CC catalyzed by the isomerase, coded by the trpC domain; the second
CC reaction is catalyzed by the synthase, coded by the trpC domain.
CC -1- CATALYTIC ACTIVITY: N-(5'-phospho-beta-D-riboseyl)-anthranilate = 1-
CC (2-carboxyphenylamino)-1-(2-carboxyphenylamino)-1-deoxy-D-ribose 5-
CC phosphate = 1-(indol-3-yl)glycerol 3-phosphate + CO(2) + H(2)O.
CC -1- PATHWAY: Tryptophan biosynthesis; third step.
CC -1- SUBUNIT: Monomer.
CC -1- SIMILARITY: In the N-terminal section; belongs to the trpC family.
CC -1- SIMILARITY: In the C-terminal section; belongs to the trpC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04960; BAA28626.1; -;
DR EMBL: AP005283; GAC00427.1; -;
DR PIR: E24723; E24723.
DR HSP; P00909; IPII.
DR HAMAP; MF_00134; fused; 1.
DR HAMAP; MF_00135; fused; 1.
DR InterPro; IPR001468; IGPS.
DR InterPro; IPR001240; PRAI.
DR Pfam; PF00218; IGPS; 1.
DR Pfam; PF00697; PRAI; 1.
DR ProDom; PD001511; IGPS; 1.
DR ProSITE; PS00614; IGPS; 1.

KW Trypophan biosynthesis; Isomerase; Lyase; Multifunctional enzyme;
 KM Decarboxylase; Complete proteome.
 FT DOMAIN 1 262 INDOL-3-GLYCEROL PHOSPHATE SYNTHASE.
 FT DOMAIN 263 474 N-(5'-PHOSPHORIBOSYL)ANTHRANILATE
 ISOMERASE.
 FT CONFLICT 88 89 SG -> AA (IN REF. 1).
 FT CONFLICT 110 110 A -> G (IN REF. 1).
 FT CONFLICT 130 131 HA -> RP (IN REF. 1).
 FT CONFLICT 153 153 A -> D (IN REF. 1).
 FT CONFLICT 302 302 L -> S (IN REF. 1).
 FT CONFLICT 343 343 D -> G (IN REF. 1).
 FT CONFLICT 381 383 MISSING (IN REF. 1).
 FT CONFLICT 454 474 AGKADGALKIKITPATISTEYH -> GWGRCRRAENFRDH
 LHIPLKV (IN REF. 1).
 SQ SEQUENCE 474 AA; 50477 MW; C347C7016BB97F9A CRC64;
 Query Match 27.4%; Score 57; DB 1; Length 474;
 Best Local Similarity 55.6%; Pred. No. 7.4;
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 13 YVEGLAGELEQARLEH 30
 DB 13 IVERGRHLEIRARLH 30

RESULT 4
 KVB3_HUMAN STANDARD; PRT; 404 AA.
 AC Q43448;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Voltage-gated potassium channel beta-3 subunit (K+ channel beta-3 subunit) (Kv-beta-3).
 GN KCNB3 OR KCNA3B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC TISSUE=Brain;
 RX MEDLINE=99074289; PubMed=9857044;
 RA Leicher T., Baehring R., Isbrandt D., Pongs O.;
 RT "Coexpression of the KCNA3B gene product with Kv1.5 leads to a novel A-type potassium channel.";
 RT J. Biol. Chem. 273:35095-35101(1998).
 CC -1- FUNCTION: Accessory potassium channel protein which modulates the activity of the pore-forming alpha subunit. Alters the functional properties of Kv1.5.
 CC -1- SUBUNIT: Forms heteromultimeric complex with alpha subunits.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- TISSUE SPECIFICITY: Brain-specific expression. Most prominent expression in cerebellum. Weaker signals detected in cortex, occipital lobe, frontal lobe and temporal lobe. Not detected in spinal cord, heart, lung, liver, kidney, pancreas, placenta and skeletal muscle.
 CC -1- DOMAIN: Alteration of functional properties of alpha subunit is mediated through N-terminal domain of beta subunit (Probable).
 CC -1- SIMILARITY: Belongs to the shaker potassium channel beta subunit family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; AF016411; AAC92499.1; -
 CC GeneW; HGNC:6230; KCNB3.
 CC MIM; 604111; -

DR GO; GO:0015459; Potassium channel regulator activity; TAS.
 DR GO; GO:000613; Potassium ion transport; TAS.
 DR InterPro; IPR001395; Aldo/ket red.
 DR InterPro; IPR005402; KCNB3 channel.
 DR InterPro; IPR005399; KCNA3 channel.
 DR Pfam; PF00248; aldo_ket_red; 1.
 DR PRINTS; PR01580; KCNB3CHANNEL.
 DR PRINTS; PR01577; KCNA3CHANNEL.
 DR PRODOM; PD000288; Aldo/ket_red; 2.
 DR TIGRfam; TIGR01293; Kv_beta; 1.
 KW Voltage-gated channel.
 KW Ion channel; Ion transport; Potassium transport;
 SQ SEQUENCE 404 AA; 43530 MW; 08265CC07929A1BA CRC64;
 Query Match 27.2%; Score 56.5; DB 1; Length 404;
 Best Local Similarity 31.1%; Pred. No. 7.2;
 Matches 14; Conservative 5; Mismatches 3; Indels 23; Gaps 2;
 QY 4 KTTGWR-----GG-----HVEGLAGELEQLR 25
 DB 145 KSKGRRSSVYTTTKIFWGQATERGJSRKHLTEGLKRSRLQ 189

RESULT 5
 SMCX_HUMAN STANDARD; PRT; 1560 AA.
 AC P41229;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Smcx protein (Xei169 protein).
 GN SMCX OR XEI169.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fibroblast;
 RX MEDLINE=94214434; PubMed=8162017;
 RA Wu J., Ellison J., Salido R., Yen P., Mohandas T., Shapiro L.J.;
 RT "Isolation and characterization of XEI169, a novel human gene that escapes X-inactivation.";
 RT Hum. Mol. Genet. 3:153-160(1994).
 RN [2]
 RP SEQUENCE OF 280-344 FROM N.A.
 RC TISSUE=Blood;
 RX MEDLINE=95038739; PubMed=7951230;
 RA Agulnik A.I., Mitchell M.J., Mattei M.-G., Borsani G., Ayner P.A., Lerner J.L., Bishop C.E.;
 RT "A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human.";
 RT Hum. Mol. Genet. 3:879-884(1994).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC Name=2;
 CC IsoId=P41229-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P41229-2; Sequence=VSP_000315;
 CC -1- TISSUE SPECIFICITY: Expressed in all tissues examined. Highest levels found in skeletal muscle.
 CC -1- MISCELLANEOUS: Escapes X-inactivation.
 CC -1- SIMILARITY: Contains 1 ARID domain.
 CC -1- SIMILARITY: Contains 1 JMW domain.
 CC -1- SIMILARITY: Contains 1 JMW domain.
 CC -1- SIMILARITY: Contains 2 PHD-type zinc fingers.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

entitles requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; 125270; AAA61302.1; -

DR EMBL; 229650; CAA82758.1; -

DR PIR; 154361; 154361.

DR Genew; HGNC:11114; SMCX.

DR MIM; 314690; -

DR InterPro; IPR001606; ARID.

DR InterPro; IPR003347; TF_JmjC.

DR InterPro; IPR003349; TF_JmjN.

DR InterPro; IPR004198; ZnF_C5HC2.

DR InterPro; IPR01965; ZnF_PHD.

DR Pfam; PF01388; ARID; 1.

DR Pfam; PF02373; JmjC; 1.

DR Pfam; PF02375; JmjN; 1.

DR Pfam; PF0628; PHD; 2.

DR Pfam; PF02928; ZF_C5HC2; 1.

DR SMART; SM00501; BRIGHT; 1.

DR SMART; SM00558; JmjC; 1.

DR SMART; SM00545; JmjN; 1.

DR SMART; SM00249; PHD; 2.

DR PROSITE; PS01359; ZF_PHD_1; 2.

DR PROSITE; PS0016; ZF_PHD_2; 1.

DR Zinc-finger; Repeat; Alternative splicing.

FT DOMAIN 13 59 JMJN.

FT DOMAIN 76 184 ARID.

FT ZN_FING 326 372 PHD-TYPE 1.

FT DOMAIN 501 617 JMJC.

FT ZN_FING 1187 1248 PHD-TYPE 2.

FT VARSPPLIC 1370 1372 Missing (in isoform 2).

FT FT /FTID=VSP 000315.

FT CONFLICT 342 342 C -> Y (IN RRP. 2).

FT SEQUENCE 1560 AA; 175804 MW; 78913F010D0CC61 CRC64;

Query Match 27.2%; Score 56.5; DB 1; Length 1560;

Best Local Similarity 28.9%; Pred. No. 31;

Matches 13; Conservative 7; Mismatches 18; Indels 7; Gaps 1;

QY 1 MEKKTGWRG-----GHVVEGLAGELQRLARLEHHPQGQREP 38

DB 1281 LTERAISWQGRARQALASDVYALIGRLAELRQLQAEPRPEPP 1325

RESULT 6

HSIV CAUCR STANDARD; PRT; 188 AA.

AC Q9A239;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE ATP-dependent protease hslv (EC 3.4.25.-).

GN HSLV OR CC3727.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;

OC Caulobacteraceae; Caulobacter.

OX NCBI_TaxID=155892;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 19089 / CB15;

RE MEDLINE=21173698; PubMed=1259647;

RA Niseman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,

RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,

RA Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,

RA DeBoy R.T., Dodson R.J., Durkin A., Stephens C., Phadke N.D., Ely B.,

RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,

RA Kohnen J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,

RA Uteberck T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,

RA Salberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

RA "Complete genome sequence of *Caulobacter crescentus*."

RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

CC -1- FUNCTION: Protease subunit of a proteasome-like degradation complex (by similarity).

CC -1- SUBUNIT: A double ring-shaped homohexamer of hslv is capped on

each side by a ring-shaped hslv homohexamer (by similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).

CC -1- SIMILARITY: Belongs to peptidase family T1B. Hslv subfamily.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AE006030; AAK25689.1; -

DR PIR; B87711; B87711.

DR HSSP; P31059; 1E94.

DR MEROPS; T01.006; -.

DR TIGR; CC3727; -.

DR HAMAP; MF_00248; -; 1.

DR InterPro; IPR001353; Peptidase_T1.

DR Pfam; PF00227; Proteasome; 1.

DR Hydrolase; Protease; Threonine protease; Complete proteome.

FT ACT_SITE 14 14 BY SIMILARITY.

FT SEQUENCE 188 AA; 19600 MW; ACCDCIFCB10BD61 CRC64;

Query Match 26.7%; Score 55.5; DB 1; Length 188;

Best Local Similarity 43.3%; Pred. No. 4.3;

Matches 13; Conservative 4; Mismatches 6; Indels 7; Gaps 1;

QY 10 GGHVVEGLAGE-----LEQRLARLEHHP 32

DB 52 GGVVAGPAGATADAFILIERLEAKLEQYP 81

RESULT 7

ALC_STRCO STANDARD; PRT; 376 AA.

AC Q9RKU4;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Putative allantoinase (EC 3.5.3.4) (Allantoinase amidinohydrolase).

GN SC06248 OR SCNH10.13 OR STRH10.13.

OS Streptomyces coelicolor.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomyces; Streptomyces.

OX NCBI_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2) / M145;

RE MEDLINE=21996410; PubMed=12000953;

RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,

RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kiese H.,

RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,

RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,

RA Huang C.-H., Kiese T., Larke L., Murphy L., Oliver K., O'Neill S.,

RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,

RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,

RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,

RA Hopwood D.A.;

RT "Complete genome sequence of the model actinomycete *Streptomyces*

RT *coelicolor* A3(2)."

RL Nature 417:141-147(2002).

CC -1- CATALYTIC ACTIVITY: Allantoin + H(2)O = (-)-ureidoglycolate +

CC urea.

CC -1- PATHWAY: Degradation of allantoin (purine catabolism); second

CC step.

CC -1- SIMILARITY: Belongs to the allantoinase family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AL939126; CAB60167.1; -

DR InterPro: IPR005164; Allantoicase.

DR Pfam: PF03561; Allantoicase; 2.

KW Hypothetical protein; Hydrolyase; Purine metabolism; Complete proteome.

SQ SEQUENCE 376 AA; 41207 MW; 2247C0B2300CA29 CRC64;

Query Match 26.0%; Score 54; DB 1; Length 376;

Best Local Similarity 44.4%; Pred. No. 14;

Matches 12; Conservative 1; Mismatches 12; Indels 2; Gaps 1;

DB 10 GGHVGLAGLELQ--LRARLEHHPOG 34

171 GGHANGFVAVSAQRTHRLKQHPDG 197

RESULT 8

KVB3 RAT STANDARD; PRT; 404 AA.

ID KVB3 RAT

AC 063454;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Voltage-gated potassium channel beta-3 subunit (K+ channel beta-3 subunit) (Kv-beta-3) (RCK beta3).

GN KCNB3 OR CKBETA3.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.

RC TISSUE=Brain cortex;

RA MEDLINE=96140552; PubMed=8549760;

RA Heinemann S.H.; Rettig J.; Wunder F.; Pongs O.;

RT "Molecular and functional characterization of a rat brain Kv beta 3 potassium channel subunit."

RL FEBS Lett. 377:383-389(1995).

CC -1- FUNCTION: Accessory potassium channel protein which modulates the activity of the pore-forming alpha subunit. Alters the functional properties of Kv1.4 but not Kv1.1 or Kv1.5.

CC -1- SUBUNIT: Forms heteromultimeric complex with alpha subunits.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).

CC TISSUE SPECIFICITY: Predominantly expressed in brain. Strongest expression in olfactory bulb and thalamic nuclei. Not detected in heart, spleen, lung, liver, skeletal muscle, kidney and testis.

CC -1- DOMAIN: Alteration of functional properties of alpha subunit is mediated through N-terminal domain of beta subunit (Probable).

CC -1- SIMILARITY: Belongs to the shaker potassium channel beta subunit family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X76723; CA54141.1; -

DR PIR: S68409; S68409.

DR InterPro: IPR001395; Aldo/ket_red.

DR InterPro: IPR005402; KCNB3 channel.

DR InterPro: IPR005399; KCNB3 channel.

DR InterPro: IPR005983; KCNB3 core.

DR Pfam: PF00248; aldo ket_red; 1.

DR PRINTS: PR01580; KCNB3CHANNEL.

DR PRINTS: PR01577; KCNB3CHANNEL.

DR ProDom: PD000288; Aldo/ket_red; 3.

DR TIGRFAMs: TIGR01293; Kv beta; 1.

KW Ionic channel; Ion transport; Potassium transport;

KW Voltage-gated channel.

SQ SEQUENCE 404 AA; 43689 MW; 7F39B80D15E71A03 CRC64;

Query Match 25.7%; Score 53.5; DB 1; Length 404;

Best Local Similarity 28.9%; Pred. No. 18;

Matches 13; Conservative 6; Mismatches 3; Indels 23; Gaps 2;

DB 4 KTTGMR-----GG-----HVEGLAGLELQLR 25

145 KKKGMRRSSVYTTTKIFWCGAETRGSLSKRHITIGLQSLDRQ 189

RESULT 9

PLSX SYNY3 STANDARD; PRT; 348 AA.

ID PLSX SYNY3

AC P73950;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Fatty acid/phospholipid synthesis protein plsx.

GN PLSX OR SLR1510.

OS Synechocystis sp. (Strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

OX NCBI_TaxID=1148;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=97061201; PubMed=8905231;

RA Kaneo T.; Sato S.; Kotani H.; Tanaka A.; Asamizu E.; Nakamura Y.,

RA Miyajima N.; Hirosewa M.; Sugiyama M.; Sasaoka S.; Kimura T.,

RA Hosouchi T.; Matsuno A.; Muraki A.; Nakazaki N.; Nario K.,

RA Okumura S.; Shimpoo S.; Takeuchi C.; Wada T.; Watanabe A.,

RA Yamada M.; Yasuda M.; Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions."

RL DNA Res. 3:109-136(1996).

CC -1- FUNCTION: Not known; probably involved in fatty acid or phospholipid synthesis (By similarity).

CC -1- SIMILARITY: Belongs to the plsx family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: D90911; BAA18017.1; -

DR PIR: S75456; S75456.

DR HAMAP: MF_00019; -; 1.

DR InterPro: IPR003664; FA_synthetis.

DR Pfam: PF02504; FA_synthetis; 1.

DR ProDom: PD006974; FA_synthetis; 1.

DR TIGRFAMs: TIGR00182; plsx; 1.

KW Fatty acid biosynthesis; phospholipid biosynthesis; Complete proteome.

SQ SEQUENCE 348 AA; 37285 MW; 12C07CC3F971A72C CRC64;

Query Match 25.5%; Score 53; DB 1; Length 348;

Best Local Similarity 34.3%; Pred. No. 18;

Matches 12; Conservative 9; Mismatches 10; Indels 4; Gaps 2;

DB 2 DEKTTGMRG--GHVVEGLAGLELQRLARLEHHPOG 34

255 EELPRGWRGKLGAIL--LAPNLKRIKQRYVDHAERG 287

RESULT 10

IFEB ASCSU STANDARD; PRT; 589 AA.

ID IFEB ASCSU

AC P23731;

DT 01-NOV-1991 (Rel. 20, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)

```

DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Intermediate filament protein B (IF-B).
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoideae).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OX Ascarididae; Ascaris.
RN (1)
RP NCB1_TaxID=6253;
RX MEDLINE=90059912; PubMed=2583097;
RA Weber K., Plesmann U., Ulrich W.;
RT "Cytoplasmic intermediate filament proteins of invertebrates are
RT closer to nuclear lamins than are vertebrate intermediate filament
RT proteins; sequence characterization of two muscle proteins of a
RT nematode."
RL EMBO J. 8:3221-3227(1989).
CC -1- SUBUNIT: A and B can form homopolymers.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: Giant body muscle cells.
CC -1- SIMILARITY: Belongs to the intermediate filament family.
DR InterPro: IPR001664; IF.
DR Pfam: PF00038; IF_tal1; IF_tal1_C.
DR PROSITE: PS00226; IF_1.
KW Intermediate filament; Coiled coil.
FT DOMAIN 1 84 HEAD.
FT FT 85 433 ROD.
FT FT 434 589 TAIL.
FT DOMAIN 85 589 TAIL.
FT FT 85 589 TAIL.
FT DOMAIN 117 130 COIL_1A.
FT DOMAIN 131 268 COIL_1B.
FT DOMAIN 269 285 LINKER_12.
FT DOMAIN 286 433 COIL_2.
SQ SEQUENCE 589 AA; 67694 MW; 5E7D9F6F0AFD6 CRC64;

Query Match 25.5%; Score 53; DB 1; Length 589;
Best Local Similarity 48.0%; Pred. No. 31;
Matches 12; Conservative 2; Mismatches 9; Indels 2; Gaps 1;

QY 3 EKTGRCGHVVEGLAGELEQLRAR 27
Db 182 EKINQWQ-HAIEDAQSELEMLRAR 204

RESULT 11
HSLV_BRUME STANDARD; PRT; 184 AA.
AC Q8YE31;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ATP-dependent protease hslv (EC 3.4.25.-).
GN HSLV OR BME12047.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCB1_TaxID=29459;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RC MEDLINE=20020109; PubMed=1156688;
RA Delvecchio V.G., Kapral V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Renik G.,
RA Jablonaki U., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elser P.H., Hagius S., O'Callaghan D., Lelsson U.-J.,
RA Haselkorn R., Kyriides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis".
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
CC complex (By similarity).
CC -1- SUBUNIT: A double ring-shaped homohexamer of hslv is capped on
each side by a ring-shaped hslv homohexamer (By similarity).
```

```

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family T1B. Hslv subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR DR EMBU; AE009637; AAL33228.1; -.
DR PIR; A13507; A13507.
DR HAMAP; MF_00248; -.
DR InterPro: IPR001353; Peptidase_T1.
DR Pfam; PF00227; proteasome; 1.
KW Hydrolyase; Protease; Threonine protease; Complete proteome.
FT ACT SITE 12 12 BY SIMILARITY.
SQ SEQUENCE 184 AA; 19839 MW; 03EDC705FDD2AA7 CRC64;

Query Match 25.2%; Score 52.5; DB 1; Length 184;
Best Local Similarity 41.4%; Pred. No. 10;
Matches 12; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

QY 11 GHVVEGLAGE-----LEQLRARLEHHP 32
Db 51 GNVIQFAGATRADATFLERLEAKLEQYF 79

RESULT 12
HSLV_BRUSU STANDARD; PRT; 184 AA.
AC O8FY11;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE ATP-dependent protease hslv (EC 3.4.25.-).
GN HSLV OR BR2080.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCB1_TaxID=29461;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=1330 / Biovar 1;
RC MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Sehnadi R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umeyam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Trettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.U., Lindler L.B., Halling S.W., Boyle C.M., Frazer C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts."
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
CC complex (By similarity).
CC -1- SUBUNIT: A double ring-shaped homohexamer of hslv is capped on
CC each side by a ring-shaped hslv homohexamer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to peptidase family T1B. Hslv subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR DR EMBU; AE014495; AAN30970.1; -.
DR TIGR; BR2080; -.
DR HAMAP; MF_00248; -. 1.
```

DR InterPro: IPR001353; Peptidase_T1.
 DR Pfam: PF00227; Proteasome_1.
 KM Hydrolyase; Protease; Threonine protease; Complete proteome.
 FT ACT SITE 12 BY SIMILARITY.
 SO SEQUENCE 184 AA; 19809 MW; 1ESBIC68F4A2FEE2 CRC64;
 Query Match 25.2%; Score 52.5; DB 1; Length 184;
 Best Local Similarity 41.4%; Pred. No. 10;
 Matches 12; Conservative 5; Mismatches 5; Indels 7; Gaps 1;
 QY 11 GHVVEGLAGE-----LEQRLARLEHHP 32
 DB 51 GNVIAGFAGATADAFLLERLEAKLEQYP 79
 RESULT 13
 XRE_BACSU STANDARD; PRT; 113 AA.
 AC P23789;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE HTH-type transcriptional regulator xre (Putative pbax repressor).
 GN XRE OR BSU12510.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RP [1]
 RX SEQUENCE FROM N.A.
 RX MEDLINE=91092505; PubMed=2125016;
 RA Wood H.E., Devine K.M., McConnell D.J.;
 RT "Characterization of a repressor gene (xre) and a
 RT temperature-sensitive allele from the Bacillus subtilis prophage,
 RT PBX-";
 RL Gene 96:83-88 (1990).
 RN [2]
 RN SEQUENCE FROM N.A.
 RN STRAIN=168 / SO113;
 RX MEDLINE=94364963; PubMed=8083174;
 RA McDonnell G.E., Wood H., Devine K.M., McConnell D.J.;
 RT "Genetic control of bacterial suicide: regulation of the induction of
 RT PBX in Bacillus subtilis."
 RL J. Bacteriol. 176:5820-5830 (1994).
 RN [3]
 RN SEQUENCE FROM N.A.
 RN STRAIN=168;
 RA Krogh S., O'Reilly M., Nolan N., Devine K.M.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RP STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunit F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borcher S.,
 RA Bortiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Comerford I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Duertnhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Hachez J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koehler P., Konigstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Maeda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Portolillo S., Prescott A.M.,
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

RA Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis."
 RL Nature 390:249-256 (1997).
 CC -1- FUNCTION: PROBABLE REPRESSOR. NECESSARY FOR THE MAINTENANCE OF
 CC THE LYSOGENIC STATE.
 CC -1- MISCELLANEOUS: THE DEFECTIVE PROPHAGE OF BACILLUS SUBTILIS 168,
 CC PBX, IS A CHROMOSOMALLY BASED ELEMENT WHICH ENCODES A
 CC NON-INFECTIONOUS PHAGE-LIKE PARTICLE WITH BACTERICIDAL ACTIVITY.
 CC PBX IS INDUCED BY AGENTS WHICH ELICIT THE SOS RESPONSE.
 CC -1- SIMILARITY: Contains 1 HTH cro/ci-type DNA-binding domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC -----
 CC EMBL: M36478; AAA22894.1; -
 CC EMBL: M36477; AAA22893.1; -
 CC EMBL: Z34287; CAA84042.1; -
 CC EMBL: Z70177; CAA94052.1; -
 CC EMBL: Z29110; CAA13108.1; -
 CC PIR: J00810; J00810.
 DR Subtilist; BG10994; xre.
 DR InterPro: IPR001387; HTH_3.
 DR Pfam: PFO1381; HTH_3; 1.
 DR SMART: SM00530; HTH_XRE; 1.
 DR PROSITE: PS50943; HTH_CROCI; 1.
 KM Transcription regulation; Repressor; DNA-binding; Complete proteome.
 FT DOMAIN 6 58
 FT DNA BIND 15 34
 FT VARIANT 4 4
 FT VARIANT 19 19
 FT VARIANT 78 78
 FT VARIANT L -> V (IN TEMPERATURE-SENSITIVE ALLELE).
 SO SEQUENCE 113 AA; 13258 MW; 23AF5C9DE3D2B31A CRC64;
 Query Match 25.0%; Score 52; DB 1; Length 113;
 Best Local Similarity 36.7%; Pred. No. 7.2;
 Matches 11; Conservative 4; Mismatches 15; Indels 0; Gaps 0;
 QY 9 RGHVVEGLAGELEQRLARLEHHPGOREP 38
 DB 10 RGRQTQEEIASHIGVSRRARYSHYENGSRP 39
 RESULT 14
 GBB2_CAEEL STANDARD; PRT; 356 AA.
 AC Q20636;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Guanine nucleotide-binding protein beta subunit 2.
 GN Gpb-2 OR F52A8.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A. AND INTERACTIONS.
 RP STRAIN=Bristol N2;
 RX MEDLINE=21231612; PubMed=11333232;
 RA van der Linden A.M., Sijmer F., Cuppen E., Plaetker R.H.A.;
 RT "The G protein beta subunit gpb-2 in Caenorhabditis elegans regulates


```

RT the G(alpha-q)alpha signaling network through interactions with
RT the regulator of G protein signaling proteins egl-10 and eat-16."
RT Genetics 158:221-235(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Briscot N2;
RA Wilkerson J.;
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Guanine nucleotide-binding proteins (G proteins) are
CC involved as a modulator or transducer in various transmembrane
CC signalling systems. The beta and gamma chains are required for the
CC GTPase activity, for replacement of GDP by GTP, and for G protein-
CC effector interaction.
CC -!- SUBUNIT: G proteins are composed of 3 units, alpha, beta and
CC gamma. Interacts with goa-1, eat-16, egl-10, egl-30.
CC -!- SIMILARITY: Contains 7 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF291847; AA055964.1; -.
DR EMBL; Z71263; CAAG5824.1; -.
DR PIR; T22478; T22478.
DR HSSP; P04901; ITBG.
DR WormSep; F52A8.2; CE17845.
DR InterPro; IPR001632; Gprotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS00082; WD_REPEATS_2; 5.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
DR Transducer; Repeat; WD repeat, Multigene family.
DR KW REPEAT 66 WD 1.
DR FT REPEAT 108 138 WD 2.
DR FT REPEAT 154 184 WD 3.
DR FT REPEAT 196 228 WD 4.
DR FT REPEAT 240 270 WD 5.
DR FT REPEAT 284 314 WD 6.
DR FT REPEAT 326 356 WD 7.
DR SQ SEQUENCE 356 AA; 39453 MW; 293AA6DCFA317D832 CRC64;

Query Match 25.0%; Score 52; DB 1; Length 356;
Best Local Similarity 40.7%; Pred. No. 24;
Matches 11; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 3 EKTGWGRGHVVEGLAGELEQLRRLRLE 29
DB 6 QPTTKGSEYVLEQLANEAELRKRLD 32

RESULT 15
AINX HUMAN STANDARD; PRT; 499 AA.
AC 016352; Q9BRCS;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alpha-Internexin (Alpha-1nx) (66 kDa neurofilament protein)
DE (Neurofilament-66) (NF-66).
GN INA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
FT

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=fetal brain;
RX MEDLINE=95287809; PubMed=7769995;
RA Chan S.-O., Chiu F.-C.;
RT "Cloning and developmental expression of human 66 kd neurofilament
RT protein."
RL Brain Res. Mol. Brain Res. 29:177-184(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=23288257; PubMed=12477932;
RA Strussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L.D., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marisina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carinci P., Prange C.,
RA Raha S.S., Igoe J., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunnarson P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.N.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: CLASS-IV NEURONAL INTERMEDIATE FILAMENT THAT IS ABLE TO
CC SELF-ASSEMBLE. IT IS INVOLVED IN THE MORPHOGENESIS OF NEURONS. IT
CC MAY FORM AN INDEPENDENT STRUCTURAL NETWORK WITHOUT THE INVOLVEMENT
CC OF OTHER NEUROFILAMENTS OR IT MAY COOPERATE WITH NF-L TO FORM THE
CC FILAMENTOUS BACKBONE TO WHICH NF-M AND NF-H ATTACH TO FORM THE
CC CROSS-BRIDGES.
CC -!- TISSUE SPECIFICITY: FOUND PREDOMINANTLY IN ADULT CNS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN BRAIN AS EARLY AS THE 16TH OF
CC GESTATION, AND INCREASED RAPIDLY AND REACHED A STEADY STATE LEVEL
CC BY THE 18TH WEEK OF GESTATION.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S78296; AA034482.1; -.
DR EMBL; BC006359; AA06359.1; -.
DR PIR; I52658; I52658.
DR Genew; HGNC:6057; INA.
DR MIM; 605338; -.
DR GO; GO:0005863; C:neurofilament; TAS.
DR GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
DR InterPro; IPR006821; Filament_head.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; Filament; 1.
DR Pfam; PF04732; Filament; 1.
DR PROSITE; PS00226; IF; 1.
DR KW Intermediate filament; Coiled coil; Neurone.
DR FT DOMAIN 1 87 HEAD.
DR FT DOMAIN 88 408 ROD.
DR FT DOMAIN 409 499 TAIL.
DR FT DOMAIN 88 129 COIL 1A.
DR FT DOMAIN 130 142 LINKER 1.
DR FT DOMAIN 143 238 COIL 1B.
DR FT DOMAIN 239 262 LINKER 2.
DR FT DOMAIN 263 408 COIL 2.

```


FT DOMAIN 449 454 POLY-GU.
 FT CONFLICT 37 41 GFRSQ -> ASVE (IN REF. 1).
 FT CONFLICT 67 67 R -> A (IN REF. 1).
 FT CONFLICT 92 92 T -> S (IN REF. 1).
 FT CONFLICT 128 132 ALRQR -> RCDT (IN REF. 1).
 FT CONFLICT 141 141 E -> Q (IN REF. 1).
 FT CONFLICT 147 152 LRDIRA -> PHLP (IN REF. 1).
 FT CONFLICT 191 198 GAERAKA -> RRRLKR (IN REF. 1).
 FT CONFLICT 244 244 A -> R (IN REF. 1).
 FT CONFLICT 263 263 S -> A (IN REF. 1).
 FT CONFLICT 301 301 S -> T (IN REF. 1).
 FT CONFLICT 310 311 EE -> DQ (IN REF. 1).
 FT CONFLICT 318 318 MISSING (IN REF. 1).
 SQ SEQUENCE 499 AA; 55390 MW; 4C972764E9E68D3E CRC64;

Query Match 25.0%; Score 52; DB 1; Length 499;
 Best Local Similarity 47.6%; Pred. No. 35;
 Matches 10; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 15 EGIAGELBOLRLRLEHHPOGO 35
 :||| |::||| | :| :
 DB 169 DGLAEVQRLRLRCEERSRGR 189

Search completed: July 15, 2004, 20:34:29
 Job time : 8.15294 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:31:20 ; Search time 10.2824 Seconds

(without alignments)
355.490 Million cell updates/sec

Title: US-10-004-381-25_COPY_1_38

Sequence: 1 MDEKTTGMRGHHVSEGLAGELRRLRLHHPQGGREP 38

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : PIR 78:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	28.4	342	2 E64905	probable sugar tra
2	59	28.4	342	2 A90894	probable transport
3	59	28.4	342	2 H85723	probable transport
4	57	27.4	474	1 E24723	trpc protein - Cor
5	57	27.4	663	2 D83545	probable helicase
6	56.5	27.2	1560	2 T54361	SMCX protein - hum
7	55.5	26.7	188	2 E87711	heat shock protein
8	54	26.0	321	2 H87683	transcription regu
9	54	26.0	1156	2 E69444	chromosome segrega
10	53.5	25.7	404	2 S68409	potassium channel
11	53.5	25.7	554	2 E87375	conserved hypochet
12	53	25.5	261	2 C75568	hypothetical prote
13	53	25.5	348	1 S75456	protein p13x - Syn
14	53	25.5	496	2 H70668	probable polyketid
15	53	25.5	589	2 S06954	intermediate filam
16	53	25.5	947	2 E86362	hypothetical prote
17	52.5	25.2	184	2 A13507	heat shock protein
18	52.5	25.2	392	2 P84650	probable protein p
19	52.5	25.2	744	2 A82822	NADH-ubiquinone ox
20	52	25.0	113	2 JQ0810	transcriptional repr
21	52	25.0	356	2 T72478	hypothetical prote
22	52	25.0	376	2 T34734	neurofilament-66 -
23	52	25.0	494	2 I52658	alpha-internexin -
24	52	25.0	504	2 I53868	alpha-internexin -
25	52	25.0	505	2 A41023	hypothetical prote
26	52	25.0	636	2 F72736	probable phosphati
27	52	25.0	2337	2 T40577	conserved hypochet
28	51.5	24.8	134	2 AH3177	cobinamide kinase
29	51.5	24.8	181	2 H64963	

30	51.5	24.8	181	2 D90977	cobinamide kinase
31	51.5	24.8	181	2 B85824	cobinamide kinase
32	51.5	24.8	273	2 G69770	manganese-contains
33	51.5	24.8	308	2 P95903	conserved hypochet
34	51.5	24.8	401	2 S66503	potassium channel
35	51.5	24.8	408	2 I59393	potassium channel
36	51.5	24.8	419	2 I55463	K+ channel beta-su
37	51.5	24.8	1156	2 A29838	paraport crystal
38	51	24.5	98	2 F90164	SSU ribosomal prot
39	51	24.5	169	2 F83372	conserved hypochet
40	51	24.5	228	2 C90250	hypothetical prote
41	51	24.5	261	2 S43218	hypothetical prote
42	51	24.5	351	2 D88065	protein T16A1.7 (1
43	51	24.5	358	2 C72643	probable 3-dehydro
44	51	24.5	429	2 F84015	maltose/maltodextr
45	51	24.5	432	2 T10894	neuronal pentraxin

ALIGNMENTS

```

RESULT 1
E64905
probable sugar transport permease protein b1514 - Escherichia coli (strain K-12)
C/Species: Escherichia coli
C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C/Accession: E64905
R/Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426617; PMID:9278503
A/Accession: E64905
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-342 <BLAT>
A/Cross-references: GB:AE000449; GB:U00096; NID:G1787790; PIDN:AC74587.1; PID:G1787793,
A/Experimental source: strain K-12, substrain MG1655
C/Superfamily: 1-arabinose transport system permease arah
C/Keywords: transmembrane protein
F/11-27/Domain: transmembrane #status predicted <TM1>
F/44-60/Domain: transmembrane #status predicted <TM2>
F/65-81/Domain: transmembrane #status predicted <TM3>
F/88-104/Domain: transmembrane #status predicted <TM4>
F/107-123/Domain: transmembrane #status predicted <TM5>
F/155-171/Domain: transmembrane #status predicted <TM6>
F/208-224/Domain: transmembrane #status predicted <TM7>
F/247-263/Domain: transmembrane #status predicted <TM8>
F/266-282/Domain: transmembrane #status predicted <TM9>
F/291-307/Domain: transmembrane #status predicted <TM10>

Query Match 28.4%; Score 59; DB 2; Length 342;
Best Local Similarity 57.1%; Pred. No. 5.4;
Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 8 WRGHHVSEGLAGELRRL 28
Db 135 WTGSKWIEGLPAELKQSLPL 155

RESULT 2
A90894
probable transport system permease protein Ece2121 [imported] - Escherichia coli (strain
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C/Accession: A90894
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gaseara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A/Reference number: A99629; MUID:21156231; PMID:11258796
A/Accession: A90894
A/Status: preliminary

```


RESULT 7

E87711
heat shock protein HslV [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
A:Accession: E87711
R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, U.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.; Kolo
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A>Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: E87711
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-188 <STO>
A:Cross-References: GB:AEO05673; NID:g13425497; PIDN:AAK25689.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC3727

Query Match 26.7% Score 55.5; DB 2; Length 188;
Best Local Similarity 43.3%; Pred. No. 8.2;
Matches 13; Conservative 4; Mismatches 6; Indels 7; Gaps 1;

Db

10 GGHVVEGLAGE-----LEQRLARLHHHP 32
||| ||| |::|::|::|
52 GGKVVAGFGAGTADAFLLERLEAKLEQYP 81

RESULT 8

H87683
transcription regulator, Arac family [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
A:Accession: H87683
R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolo
n, U.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.; Kolo
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A>Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: H87683
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-321 <STO>
A:Cross-References: GB:AEO05673; NID:g13425234; PIDN:AAK25468.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC3506

Query Match 26.0% Score 54; DB 2; Length 321;
Best Local Similarity 30.3%; Pred. No. 23;
Matches 10; Conservative 7; Mismatches 16; Indels 0; Gaps 0;

Db

2 DEKTGWRGHVEGLAGELEQLRARLEHHPG 34
| | | | |::|::|::|
205 DPDRVWMAAQILRGAGTLDRLARIRIDLAPRG 237

RESULT 9

B69444
chromosome segregation protein (smc1) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Jun-2000
A:Accession: B69444
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Moose, C.R.; Venter, J.C.
A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343; PMID:9389475

A.Accession: B69444
A.Status: Preliminary; nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-1156 <KLE>
A.Cross-references: GB:AE000995; GB:AE00782; NID:g2669318; PIDN:AAB89690.1; PID:g26490C
C.Superfamily: chromosome segregation protein SMC1

Query Match 26.0%; Score 54; DB 2; Length 1156;
Best Local Similarity 44.4%; Pred. No. 89;
Matches 12; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

DB 349 GEYDELMANKELLORLEEVDKKHRE 375

RESULT 10
68409 potassium channel Kv beta-3 chain - rat
C.Species: Rattus norvegicus (Norway rat)
C.Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C.Accession: S68409; S72562
R.Heinemann, S.H.; Rettig, J.; Wunder, F.; Pongs, O.
FEBS Lett. 377, 383-389, 1995
A.Title: Molecular and functional characterization of a rat brain K(v)-beta3 potassium C
A.Reference number: S68409; MUID:96140552; PMID:8549760
A.Accession: S68409
A>Status: nucleic acid sequence not shown
A.Molecule type: mRNA
A.Residues: 1-404 <HEI>
A.Cross-references: EMBL:X76723
A.Experimental source: cortex
R.Rettig, J.
submitted to the EMBL Data Library, December 1993
A.Reference number: S72562
A.Accession: S72562
A.Molecule type: mRNA
A.Residues: 1-245, 'P', 247-324, 'K', 326-404 <RET>
A.Cross-references: EMBL:X76723; NID:g1246833; PIDN:CMA54141.1; PID:g1246834
A.Experimental source: cortex
C.Genetics:
A.Gene: RCKbeta3
C.Superfamily: fission yeast pyridoxine 4-dehydrogenase
C.Keywords: voltage-gated ion channel

Query Match 25.7%; Score 53.5; DB 2; Length 404;
Best Local Similarity 28.9%; Pred. No. 34;
Matches 13; Conservative 6; Mismatches 3; Indels 23; Gaps 2;

OY 4 KTTGMR-----GG-----HVVEGLAGLEQLR 25
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 145 KSKCMRRSSVITTKIFMGQALERTGLSRKHIIIEGLQGSLDRLO 189

RESULT 11
E87375 conserved hypothetical protein CC1017 [imported] - Caulobacter crescentus
C.Species: Caulobacter crescentus
C.Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C.Accession: E87375
R.Nielsen, W.C.; Felblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
M.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A>Title: Complete Genome Sequence of Caulobacter crescentus.
A.Reference number: A87249; MUID:21173696; PMID:11259647
A.Accession: E87375
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-554 <STO>
A.Cross-references: GB:AE005673; NID:g13422309; PIDN:AAK23001.1; GSFPDB:GN00148
C.Genetics:
A.Gene: CC1017

```

Query Match Similarity      25.7%; Score 53.5; DB 2; Length 554;
Best Local Similarity     35.9%; Pred. No. 47;
Matches    14; Conservative   4; Mismatches    16; Indels       5; Gaps        1;

OY      5 TTGWRGHVVE-----GLAGLEIOLRLARLHHNPGQREP 38
         |||::|||::|||::|||::|||::|||::|||::|||
Db       45 STEWSAGYVTDVNYTFGYGEINPLRCRLPLTLVGRRHAP 83


RESULT 12
C75568
hypothetical protein - Deinococcus radiodurans (strain R1)
CISpecies: Deinococcus radiodurans
CIDate: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
CAccession: C75568
R.White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; PMID:20036896; PMID:10567266
A:Accession: C75568
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-261 <WH>
A:Cross-references: GB:AEO01867; GB:AEO00513; NID:G6457693; PID:NABF09637.1; PID:G6457707
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0043
A:Map position: 1


Query Match          25.5%; Score 53; DB 2; Length 261;
Best Local Similarity 32.6%; Pred. No. 25;
Matches    15; Conservative   5; Mismatches    14; Indels    12; Gaps     1;

OY      2 DEKTTGWR-----GHNVEGIAGLEIOLRLARLHHNPGQ 35
         |||::|||::|||::|||::|||::|||::|||::|||
Db       216 DEERAPWRLALRDITLTAYLGGRGVRTDLARESEAFVYLKKHPREGQ 261


RESULT 13
S75456
protein plex - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein plex1510
CISpecies: Synechocystis sp.
A:Variety: PCC 6803
CIDate: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2003
CAccession: S75456
R.Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Kikuchi, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Matsunabe, A.; Yamada, M.; Yasuda, D.N.A. Res. 3, 109-116, 1996
A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S75456
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-948 <KAN>
A:Cross-references: EMBL:D90911; GB:AEO01339; NID:g1653083; PID:BAA18017.1; PID:g165310
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: Plax
C:Superfamily: phospholipid biosynthesis protein, plxk type


Query Match          25.5%; Score 53; DB 1; Length 348;
Best Local Similarity 34.3%; Pred. No. 33;
Matches    12; Conservative   9; Mismatches    10; Indels     4; Gaps     2;

OY      2 DEKTTGWR--GNHVEGIAGLEIOLRLARLHHNPGQ 34
         ::|||::|||::|||::|||::|||::|||::|||
Db       255 BELPRGMKGKLGAII--LARNLKRIKORVRDADENG 287
```

```

RESULT 14
H70668
probable polyketide synthase - Mycobacterium tuberculosis (strain H37Rv)
C|Species: Mycobacterium tuberculosis
C|Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C|Accession: H70668; S73074
R|Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
R|Comor, R.A.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
R|Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A|Authors: Squires, R.; Sulistiono, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A|Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome :
A|Reference number: A70500; NWID:98295987; PMID:9634230
A|Accession: H70668
A|Status: preliminary; nucleic acid sequence not shown; translation not shown
A|Molecule type: DNA
A|Residues: 1-496 <COL>
A|Cross-references: GB:I283858; GB:A1123456; NID:g560506; PIDN:CAB06102.1; PID:g1781166
A|Experimental source: strain H37Rv
R|Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, September 1994
A|Description: Mycobacterium tuberculosis cosmid tbc2.
A|Reference number: S73053
A|Accession: S73074
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-273,275-496 <SMI>
A|Cross-references: EMBL:U00024; NID:g560506; PIDN:AAA50929.1; PID:g560508
C|Genetics:
A|Gene: PKB15
A|Start codon: GTG
C|Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I; 3-oxoacyl-[acyl-carrier-prot-
F|68-467/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>

Query Match      25.5%; Score 53; DB 2; Length 496;
Best Local Similarity 40.7%; Pred. No. 49;
Matches 11; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY      12 HVVEGIAGLEQLRLRNHPQGOREP 38
       |: : | ||| :||| :| :|
DB      22 HYLKVAVELDETRARLRREYQRATRP 48

RESULT 15
S06954
Intermediate filament protein B - common roundworm
C|Species: Ascaris lumbricoides (common roundworm)
C|Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 13-Sep-1998
C|Accession: S06954
R|Weber, K.; Plessmann, U.; Ulrich, W.
EMBO J. 8, 3221-3227, 1989
A|Title: Cytoplasmic intermediate filament proteins of invertebrates are closer to nucle-
of a nematode.
A|Reference number: S06954; NWID:90059912; PMID:2583097
A|Accession: S06954
A|Molecule type: protein
A|Residues: 1-589 <WEB>
C|Superfamily: Intermediate filament protein Av71
C|Keywords: coiled coil

Query Match      25.5%; Score 53; DB 2; Length 589;
Best Local Similarity 48.0%; Pred. No. 59;
Matches 12; Conservative 2; Mismatches 9; Indels 2; Gaps 1;

QY      3 EXTTGMRGSHVVEGLAELEQLRAR 27
       ||| :| :| :| ||| |||
DB      182 EKINQWQ--HAIEDAQSELEMLRAR 204

Search completed: July 15, 2004, 20:36:21
Job time : 11.2824 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 20:32:20 ; Search time 12.9647 Seconds
(without alignments)
151.318 Million cell updates/sec

Title: US-10-004-381-25_COPY_1_38

Perfect score: 208

Sequence: 1 MDEKTGWRGVHVEGLAGLEQLRLARLEHHHPOGQREP 38

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	30.3	543	3	US-09-413-814-91 Sequence 91, Appl
2	63	30.3	544	3	US-09-413-814-81 Sequence 81, Appl
3	59.5	28.6	543	4	US-09-252-991A-29830 Sequence 29830, A
4	58.5	28.1	197	4	US-09-252-991A-24825 Sequence 24825, A
5	58	27.9	113	2	US-08-248-839C-6 Sequence 6, Appl
6	58	27.9	628	4	US-09-252-991A-22131 Sequence 22131, A
7	57.5	27.6	239	4	US-09-134-000C-5025 Sequence 5025, Ap
8	57	27.4	356	4	US-09-252-991A-19549 Sequence 19549, A
9	57	27.4	904	4	US-09-252-991A-29119 Sequence 29119, A
10	55.5	26.7	1388	4	US-09-572-191-2 Sequence 2, Appl
11	55.5	26.7	1388	4	US-09-723-262-2 Sequence 2, Appl
12	55.5	26.7	1388	4	US-09-723-219-2 Sequence 2, Appl
13	54	26.0	185	4	US-09-252-991A-19647 Sequence 19647, A
14	53.5	25.7	266	4	US-09-252-991A-24870 Sequence 24870, A
15	53.5	25.7	622	4	US-09-252-991A-19069 Sequence 19069, A
16	53	25.5	152	4	US-09-252-991A-18469 Sequence 18469, A
17	53	25.5	485	4	US-08-311-731A-3 Sequence 3, Appl
18	52.5	25.2	69	4	US-09-328-352-4132 Sequence 4132, Ap
19	52.5	25.2	220	4	US-09-252-991A-24796 Sequence 24796, A
20	52.5	25.2	279	4	US-09-252-991A-19165 Sequence 19165, A
21	52.5	25.2	293	4	US-09-252-991A-20071 Sequence 20071, A
22	52.5	25.2	459	4	US-09-134-000C-6639 Sequence 6639, Ap
23	52.5	25.2	524	4	US-09-252-991A-19671 Sequence 19671, A
24	52.5	25.2	631	4	US-09-252-991A-26444 Sequence 26444, A
25	52	25.0	113	2	US-08-248-839C-2 Sequence 2, Appl
26	52	25.0	113	2	US-08-248-839C-8 Sequence 8, Appl
27	52	25.0	152	4	US-09-252-991A-31405 Sequence 31405, A

28	52	25.0	261	4	US-09-252-991A-28545 Sequence 28545, A
29	52	25.0	423	2	US-08-715-554-2 Sequence 2, Appl
30	52	25.0	423	2	US-08-583-118-2 Sequence 2, Appl
31	52	25.0	499	4	US-09-252-991A-27221 Sequence 27221, A
32	52	25.0	741	4	US-09-252-991A-27062 Sequence 27062, A
33	51.5	24.8	329	2	US-08-606-143-1 Sequence 1, Appl
34	51.5	24.8	329	2	US-08-606-143-3 Sequence 3, Appl
35	51.5	24.8	329	2	US-09-328-352-5059 Sequence 5059, Ap
36	51.5	24.8	166	4	US-09-252-991A-24661 Sequence 24661, A
37	51	24.5	179	4	US-09-252-991A-30679 Sequence 30679, A
38	51	24.5	293	4	US-09-252-991A-27745 Sequence 27745, A
39	51	24.5	293	4	US-09-252-991A-28815 Sequence 28815, A
40	51	24.5	324	4	US-09-489-039A-10737 Sequence 10737, A
41	51	24.5	427	4	US-09-252-991A-20199 Sequence 20199, A
42	51	24.5	428	4	US-08-631-607-7 Sequence 7, Appl
43	51	24.5	430	1	US-09-098-358B-7 Sequence 7, Appl
44	51	24.5	430	1	US-09-098-358B-7 Sequence 7, Appl
45	51	24.5	430	1	US-09-098-358B-7 Sequence 7, Appl

ALIGNMENTS

```
RESULT 1
US-09-413-814-91
; Sequence 91, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Bayer, Stefan
; APPLICANT: Bioeker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413, 814
; EARLIER FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 91
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-91

Query Match
Best Local Similarity 30.1%; Score 63; DB 3; Length 543;
Matches 13; Conservative 6; Mismatches 10; Indels 8; Gaps 1;

QY 8 WRGVHVEG-----LAGELEQLRLARLEHHHPOGQ 36
Db 235 WLSGEVLSEGLKVTYTKLSGALRRARVPVDHDPARR 271

RESULT 2
US-09-413-814-81
; Sequence 81, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Bayer, Stefan
; APPLICANT: Bioeker, Helmut
; APPLICANT: Brandt, Petra
```

APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hoffe, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
EARLIER FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 107
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 81
LENGTH: 544
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-81

Query Match 30.3%; Score 63; DB 3; Length 544;
Best Local Similarity 35.1%; Pred. No. 0.97;
Matches 13; Conservative 6; Mismatches 10; Indels 8; Gaps 1;

QY 8 WRGHHVEG-----LAGELEQLRRLHHHPOGOR 36
DB 236 WLSEVLEGELERLYTKLSGALRRARVPDHPAGRR 272

RESULT 3
US-09-252-991A-29830
Sequence 29830, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 29830
LENGTH: 543
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29830

Query Match 28.6%; Score 59.5; DB 4; Length 543;
Best Local Similarity 42.9%; Pred. No. 3;
Matches 15; Conservative 3; Mismatches 14; Indels 3; Gaps 2;

QY 7 GRRGGHVVETGLAG--ELLEQLRLRLHHHPOGOR 38
DB 183 GQSGAHAFGLADLVLELDLGVAGQRHRRPRDWRHP 217

RESULT 4
US-09-252-991A-24825
Sequence 24825, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 24825
LENGTH: 197
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24825

Query Match 28.1%; Score 58.5; DB 4; Length 197;
Best Local Similarity 44.8%; Pred. No. 1.3;
Matches 13; Conservative 2; Mismatches 13; Indels 1; Gaps 1;

QY 8 WRGHHVEGLAGLEQLRRL-ELHHPOGO 35
DB 112 WRGHAEPDGPCHRRRRLPLRRHHPOGR 140

RESULT 5
US-08-248-839C-6
Sequence 6, Application US/08248839C
Patent No. 5843702
GENERAL INFORMATION:
APPLICANT: McConnell, David
APPLICANT: Devine, Kevin
APPLICANT: O'Kane, Charles
TITLE OF INVENTION: A Gene Expression System
NUMBER OF SEQUENCES: 185
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58437020 No. 5843702disk of No. 5843702th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/248,839C
FILING DATE: 25-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Grege, Valeta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 3614.214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 113 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-248-839C-6

Query Match 27.9%; Score 58; DB 2; Length 113;
Best Local Similarity 40.0%; Pred. No. 0.81;
Matches 12; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY 9 RGHVVEGLAGLEQLRRLHHHPOGOR 38
DB 10 RGRQVEIAGHIGVSRARYSHYENGRESE 39

RESULT 6
US-09-252-991A-22131
Sequence 22131, Application US/09252991A
Patent No. 6551795


```
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ PRIOR FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 22131
/ LENGTH: 628
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22131

Query Match      27.4%; Score 58; DB 4; Length 628;
Best Local Similarity 44.4%; Pred. No. 5.8;
Matches 16; Conservative 1; Mismatches 11; Indels 8; Gaps 1;

QY      9 RCGHVE-----GLAGEQLRLARLHHHPGQR 36
Db      126 RCGHAEVVVGLAVGGGLAVGLAVPGRGADHPVGR 161

RESULT 7
US-09-134-000C-5025
/ Sequence 5025, Application US/09134000C
/ Patent No. 6617156
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Scamm et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
/ TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 032796-032
/ CURRENT APPLICATION NUMBER: US/09/134,000C
/ CURRENT FILING DATE: 1998-08-13
/ PRIOR APPLICATION NUMBER: US 60/055,778
/ PRIOR FILING DATE: 1997-08-15
/ NUMBER OF SEQ ID NOS: 6812
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 5025
/ LENGTH: 239
/ TYPE: PRT
/ ORGANISM: Enterococcus faecalis
US-09-134-000C-5025

Query Match      27.6%; Score 57.5; DB 4; Length 239;
Best Local Similarity 52.2%; Pred. No. 2.3;
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY      8 WRGHHVEGLAGEQLRLARLHH 30
Db      34 W-GEVITGIMASMEQERAKEH 55

RESULT 8
US-09-252-991A-19549
/ Sequence 19549, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142

/ SEQ ID NO 19549
/ LENGTH: 356
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19549

Query Match      27.4%; Score 57; DB 4; Length 356;
Best Local Similarity 45.5%; Pred. No. 4.2;
Matches 15; Conservative 3; Mismatches 9; Indels 6; Gaps 2;

QY      7 GWRGHHVEGLAGEQLRLARLHHHPQ--GQRE 37
Db      183 GRRGGHAGGEPQ-----RATGHPHVAVGGRQ 211

RESULT 9
US-09-252-991A-29119
/ Sequence 29119, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 29119
/ LENGTH: 904
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29119

Query Match      27.4%; Score 57; DB 4; Length 904;
Best Local Similarity 36.1%; Pred. No. 12;
Matches 13; Conservative 3; Mismatches 20; Indels 0; Gaps 0;

QY      1 MDKTTGWRGHHVEGLAGEQLRLARLHHHPGQR 36
Db      395 LDLEPLAWGCGDNLDPARLRLSFQARPRATPPQGR 430

RESULT 10
US-09-572-191-2
/ Sequence 2, Application US/09572191
/ Patent No. 6355466
/ GENERAL INFORMATION:
/ APPLICANT: Bernard, Christophe
/ APPLICANT: Sakowicz, Roman
/ APPLICANT: Wood, Kenneth
/ TITLE OF INVENTION: No. 6355466 motor proteins and methods for
/ TITLE OF INVENTION: their use
/ FILE REFERENCE: 1017
/ CURRENT APPLICATION NUMBER: US/09/572,191
/ CURRENT FILING DATE: 2000-05-17
/ NUMBER OF SEQ ID NOS: 6
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2
/ LENGTH: 1388
/ TYPE: PRT
/ ORGANISM: Human
US-09-572-191-2

Query Match      26.7%; Score 55.5; DB 4; Length 1388;
Best Local Similarity 37.5%; Pred. No. 32;
Matches 12; Conservative 6; Mismatches 7; Indels 7; Gaps 1;

QY      9 RCGHVE-----GLAGEQLRLARLHHHPQ 33
```


; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 19069
 ; LENGTH: 692
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-19069

Query Match 25.7%; Score 53.5; DB 4; Length 692;
 Best Local Similarity 42.9%; Pred. No. 28;
 Matches 12; Conservative 4; Mismatches 9; Indels 3; Gaps 1;

QY 6 TGRGGHVVEGLAGELQLELRLHHPQ 33
 DB 287 TGRHVGSEYIEGVSGLQRL--AERHPR 311

Search completed: July 15, 2004, 20:37:00
 Job time : 12.9647 secs

This Page Blank (uspto)

This Page Blank (uspto)